

Рассмотрев переводческие лексические трансформации в исследуемых текстах, мы пришли к выводу, что чаще всего при переводе англоязычных официально-деловых текстов используются словарные переводные эквиваленты. Иногда встречаются генерализация, конкретизация, замена частей речи и смысловое развитие. При автоматическом переводе с помощью программы машинного перевода наиболее характерными для данных текстов являются адекватная замена и лексическое развитие, правда встречается они редко. Особую трудность в системе машинного перевода вызывают многозначные слова (*people, food, water*), значение которых зависит от контекста и может значительно различаться. Например, лексическая единица *water* переводится как ‘водный ресурс’, ‘водоснабжение’.

ЛИТЕРАТУРА

1. *Комиссаров, В. Н.* Современное переводоведение : курс лекций / В. Н. Комиссаров. – М. : ЭТС, 1999. – 192 с.
2. *Бархударов, Л. С.* Язык и перевод. Вопросы общей и частной теории перевода / Л. С. Бархударов. – 2-е изд. – М. : ЛКИ, 2007. – 240 с.

In the article the results and main conclusions of the research of lexical transformations in automatic translation of English-language texts of 17 UN goals for sustainable development into Russian are presented.

А. В. Скрябина
Минск, МГЛУ

АВТОМАТИЧЕСКОЕ ИНДЕКСИРОВАНИЕ АНГЛОЯЗЫЧНЫХ ПУБЛИЦИСТИЧЕСКИХ ТЕКСТОВ

В статье рассматривается проблематика автоматического индексирования англоязычных новостных текстов на основе выделения ключевых слов.

В настоящее время системы, автоматизирующие обработку текста с целью выявления его смысла и структуры, крайне немногочисленны или находятся в экспериментальной стадии разработки. В основном такие системы выделяют лишь смысл отдельных слов либо словосочетаний, но структура целого предложения целиком не анализируется. Сам процесс описания смыслового содержания текста средствами информационно-поискового языка (ИПЯ) называется индексированием. А именно это процесс присвоения документам и запросам или их частям индексов. Итоговая продукция этих процессов содержится в поисковом образе документа.

В нашем исследовании материалом послужили англоязычные новостные тексты по теме «Культура». Все публицистические тексты взяты с интернет-ресурсов www.bbc.com и www.theguardian.com. В нашей работе устанавли-

вается основное содержание каждого из отобранных текстов. Нами были выделены ключевые слова по конкретным тематикам, которые влияют на понимание всего текста. Ключевые слова в нашей работе – это знаменательные лексические единицы текста, концентрированно представляющие его основное содержание и обладающие наибольшей частотой встречаемости.

Существует несколько критериев выделения ключевых слов [1]:

- 1) статистический критерий, или критерий частоты;
- 2) критерий встречаемости слов (критерий повторяемости ключевых слов в тексте);
- 3) критерий вхождения слова в первое предложение текста, в заголовок, в реферат и т.п.;
- 4) словарный критерий (на основе наличия или отсутствия ключевого слова в заданном списке слов);
- 5) синтаксический критерий (на основе синтаксических отношений ключевых слов, его позиции среди других лексических единиц; разновидностями этого критерия являются критерий активности и критерий воздействия);
- 6) диалоговый критерий (когда выбор ключевых слов осуществляется человеком из числа предлагаемых компьютером);
- 7) анкетный, или матричный критерий (заранее задается схема содержания текста, сама структура которой предполагает выделение типов определенных ключевых слов);
- 8) критерий протяженности, учитывающий количество предложений между первым и последним упоминанием данного предмета речи;
- 9) интуитивный критерий (выбор осуществляется испытуемыми на основе субъективного восприятия);
- 10) контекстологический критерий (предполагает учет денотативно тождественных лексических единиц, поскольку ключевое слово обязательно совпадает в тексте только с одной лексической единицей, им может быть и устойчивое словосочетание, и различные синонимы, и местоименные замены).

В работе мы использовали статистический метод. Процесс индексирования включает несколько главных и второстепенных операций, которые выполняются последовательно или одновременно [2].

1. Анализ содержания документа и выбор из текста номинативных лексических единиц, существенных с точки зрения его содержания.

2. Формирование перечня ключевых слов (морфологически нормализованных лексических единиц), используемых в процессе свободного координатного индексирования.

3. Избыточное индексирование – введение дополнительных лексических единиц, связанных по смыслу с исходными ключевыми словами.

Индексирование – одно из основных понятий информационного поиска. Цель данного процесса в информационной системе – приписать каждому документу и запросу некоторое множество «индексов» – идентификаторов,

отражающих содержание документа. Идентификаторы называют рубриками, ключевыми словами, дескрипторами, в зависимости от типа используемого ИПЯ. Эти «индексы» отражают содержание документа и управляют поиском, приводя к тем документам, термины которых оказываются наиболее сходными с терминами запроса.

Всего для анализа нами было отобрано 40 англоязычных новостных публицистических текстов по теме «Культура», и мы выделили более узкие тематики в данной области. Все текстовые единицы по анализируемой тематике были разделены на десять более узких групп на основании тематической принадлежности публицистических текстов. В созданной нами базе данных поисковый образ, например, для тематики «Books» выглядит следующим образом, что представлено в таблице.

Поисковый образ для тематики «Books»

Ключевое слово	Грамматические формы ключевых слов		
AUTHOR	authors		
BESTSELLING			
BOOK	books		
CRIME	crimes		
DETECTIVE	detectives		
FICTION	fictions		
NOVEL	novels		
NOVELIST	novelists		
PRIZE	prizes		
READ	reads	reading	read
SPONSORSHIP	sponsorships		
TOPIC	topics		
WRITE	writes	writing	wrote
WRITER	writers		
YEAR	years		

На основании разработанных поисковых образов по каждой тематике был составлен алгоритм автоматического индексирования англоязычных новостных текстов.

ЛИТЕРАТУРА

1. Search pattern of a document [Электронный ресурс]. – Режим доступа : <http://www.finam.ru/dictionary/wordf0229200013/default.asp?n=1>. – Дата доступа : 12.04.2017.

2. *Захаров, В. П.* Информационные системы (документальный поиск) / В. П. Захаров. – СПб., 2002. – 188 с.

The article analyses the procedure of creating a search image for English publicist texts, which is based on the algorithm of identifying key or prop words.

А. Л. Холод
Минск, МГЛУ

ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ СИСТЕМЫ АВТОМАТИЧЕСКОГО ОБНАРУЖЕНИЯ СРЕДСТВ ВЫРАЖЕНИЯ СТРАТЕГИИ «ФЛЕЙМ» В ВИРТУАЛЬНОЙ КОММУНИКАЦИИ

В статье рассматриваются десять основных классов лексических единиц, однозначно указывающих на наличие в испаноязычных твитах деструктивной стратегии «флейм». Отмечается, что деструктивный эффект высказывания может быть усилен путем использования особых идеографических знаков (эмодзи), передающих в виртуальной коммуникации определенную эмоцию или эмоциональное состояние пользователя. В работе также приводятся второстепенные маркеры реализации флейма, которые проявляют свой деструктивный потенциал в случае, если их количество в твите превышает две единицы. Выявленные в ходе исследования маркеры составили лингвистическую базу данных системы автоматического обнаружения и блокировки твитов с флеймом.

Периодическое отклонение от норм и правил кооперативного общения в Интернете, а также агрессивное речевое поведение его пользователей явно указывает на деструктивный характер виртуальной коммуникации, в рамках которой можно выделить различные деструктивные стратегии, и в частности флейм. Флейм представляет собой эмоциональное обсуждение определенной темы, сопровождающееся нарушением принципов конструктивной дискуссии. Флейм относится к запрещенным сетевым коммуникативным стратегиям, однако особенности идентификации пользователей превращают контроль за соблюдением правил сетевого общения в довольно сложную задачу. Модераторы блогов, форумов, социальных сетей и других сетевых средств общения блокируют сообщения агрессивных пользователей вручную. В то же время выявленные в полном объеме вербальные и невербальные средства реализации данной стратегии могут быть положены в основу создания системы автоматического обнаружения и блокировки сообщений с флеймом.

Материалом исследования послужили 278 испаноязычных аккаунтов социальной сети Твиттер (общим объемом более 250 000 твитов), из которых детально было проанализировано 5000 твитов и выделено 720 твитов, содержащих признаки деструктивной стратегии «флейм». Общение на этих аккаунтах характеризуется повышенной эмоциональностью и экспрессивностью, а также категоричностью высказываемых точек зрения. На первом этапе исследования отобранных твитов были определены основные классы вербальных и невербальных маркеров, однозначно указывающих на наличие деструктивной стратегии, важных для разработки процедуры автоматического