

4. Макарова Л.С., Долуденко Е.А. Материалы по курсу теории и практики перевода.[Электронный ресурс].
5. Левицкая Т. Р, Фитерман А. М. Проблемы перевода. 1976.

Оригиналы статей

1. Интернет-адрес: <http://www.yourbbsucks.com/forum/showthread.php?t=494>.
2. Интернет-адрес: http://www.cbsnews.com/8301-18563_162-688311.html.
3. Интернет-адрес: <http://www.bbc.co.uk/news/world-20233941>.
4. Интернет-адрес:
http://www.bbc.co.uk/russian/international/2012/11/121103_usa2012_obama_romney_elections.shtml.
5. Интернет-адрес: http://www.ideasinactiontv.com/tcs_daily/2005/02/greenwashing-putin.html.
6. Интернет-адрес: <http://www.guardian.co.uk/world/2005/jan/14/usa.guantanamo>.
7. Интернет-адрес: <http://www.pbs.org/wgbh/americanexperience/features/primary-resources/lbj-union67/>.
8. Интернет-адрес:
<http://webcache.googleusercontent.com/search?q=cache:taGBytCc7TMJ:www.usnews.com/usnews/opinion/articles/060130/30edit.htm+&cd=2&hl=ru&ct=clnk&gl=by>.
9. Интернет-адрес: <http://www.pbs.org/wgbh/americanexperience/features/primary-resources/lbj-union67/>.

ОШИБКИ НА СЕМАНТИЧЕСКОМ УРОВНЕ В СИСТЕМАХ МАШИННОГО ПЕРЕВОДА: ПРИЧИНЫ ВОЗНИКНОВЕНИЯ И СПОСОБЫ УСТРАНЕНИЯ

Т. И. Окрут

При общении человека и ЭВМ всегда возникал семантический барьер, и работа любой системы автоматической обработки естественного языка подразумевает под собой наличие семантических ошибок, в том числе и работа систем машинного перевода (МП).

Для проведения анализа семантических ошибок в системах машинного перевода сначала необходимо разобраться в том, как эти системы устроены.

В современных системах машинного перевода можно выделить системы, основанные на правилах, на примерах и статистические системы МП [1, с. 110].

Системы МП на правилах создаются на основе лингвистической информации об исходном и переводном языках. Они состоят из двуязычных словарей и грамматик, охватывающих основные семантические, морфологические, синтаксические закономерности каждого языка. Как правило, эти системы используют формальный язык-посредник в качестве основы для создания текста на языке перевода [2, с. 179].

В рамках метода, основанного на правилах, выделяют интерлингвистические и трансферные системы машинного перевода.

Среди представителей систем машинного перевода на правилах можно выделить продукты компаний PROMT, ImTranslator, Systran (трансферные системы) и системы машинного перевода семейства ЭТАП (интерлингвистические системы) [3].

В системах МП на примерах используется двуязычные корпуса текстов, составленные из пар предложений. Основная идея: при накоплении достаточно большого объема ранее переведенных текстов велика вероятность того, что большая часть последующих текстов будет аналогична уже переведенным вручную.

К системам МП на примерах относят системы на базе технологии «памяти переводов» (Translation Memory) [3, с. 114]. «В общем массиве текста система находит сегменты, которые уже были однажды переведены, и берет перевод из базы переводов Translation Memory» [4].

Статистические системы МП основаны на поиске наиболее вероятного перевода предложения с использованием данных, полученных из двуязычной совокупности текстов.

Наиболее яркими представителями данной технологии являются системы компаний Google и Яндекс.

Рассмотрим строение системы МП компании Яндекс. Она состоит из модели перевода, модели языка и декодера. В модели перевода подбираются параллельные документы, потом в них – пары предложений, слов и словосочетаний. Для создания модели языка система изучает тексты на нужном языке и составляет список всех употребленных в них слов и словосочетаний с указанием частоты их использования. Непосредственно переводом занимается декодер. Он подбирает все варианты перевода для предложений исходного текста, сочетая между собой фразы из модели перевода, и сортирует их по убыванию вероятности [5].

Несмотря на многообразие вышеописанных систем, их всех объединяет общая тенденция: возникают ошибки, и в первую очередь нас интересуют ошибки на семантическом уровне.

В качестве объектов для исследования было выбрано 3 представителя систем машинного перевода: онлайн переводчики компаний Promt (трансфертная СМП), Yandex (статистическая СМП) и система машинного перевода ЭТАП (интерлингвистическая СМП). На примере работы этих систем в рамках русско-английского перевода изучены семантические ошибки, а также рассмотрены возможные варианты их предотвращения.

В качестве текста для перевода был взят отрывок из произведения «Преступление и наказание» Ф.М. Достоевского. В качестве челове-

ского перевода был взят соответствующий отрывок из работы Констанс Гарнетт (табл. 1).

Таблица 1

Вариант человеческого перевода от Констанс Гарнетт

Ф.М. Достоевский «Преступление и наказание»	Перевод от Констанс Гарнетт
«Я так и знал! — бормотал он в смущении, — я так и думал! Это уж всего сквернее! Вот эдакая какая-нибудь глупость, какая-нибудь пошлейшая мелочь, весь замысел может испортить! Да, слишком приметная шляпа... Смешная, потому и приметная... К моим лохмотьям непременно нужна фуражка, хотя бы старый блин какой-нибудь, а не этот урод. Никто таких не носит, за версту заметят, запомнят... главное, потом запомнят, ан и улика...»	"I knew it," he muttered in confusion, "I thought so! That's the worst of all! Why, a stupid thing like this, the most trivial detail might spoil the whole plan. Yes, my hat is too noticeable. . . . It looks absurd and that makes it noticeable. . . . With my rags I ought to wear a cap, any sort of old pancake, but not this grotesque thing. Nobody wears such a hat, it would be noticed a mile off, it would be remembered. . . . What matters is that people would remember it, and that would give them a clue. . ."

Далее следуют варианты систем машинного перевода (табл. 2).

Появление семантических ошибок в первую очередь обусловлено омонимией (*ан* как союз и сокращение) и полисемией слов (например, слова *мелочь*, *глупость*, *улика*). Также системы Promt и Yandex не смогли распознать превосходную степень слова *пошлый* по причине отсутствия полной морфологической информации о слове (в случае с Promt) и из-за отсутствия в модели перевода фраз со словом *пошлейший* (в случае с Yandex). Устойчивое выражение *за версту* система Yandex не смогла распознать в контексте, в то время как для обособленного выражения перевод дается верный. Система МП Promt предложила нефразеологический вариант *far off*, а система ЭТАП дала пословный перевод за неимением эквивалента в электронном словаре.

Таким образом, качество систем МП на правилах зависит от объема словаря и объема информации, приписываемой лексическим единицам, от тщательности проектирования алгоритмов анализа и синтеза, а также от параметров программного обеспечения. Статистические же системы МП требуют постоянного пополнения параллельных корпусов. При этом все этапы машинного перевода взаимосвязаны, ошибки на одном из этапов влекут за собой появление ошибок на других этапах.

В зависимости от систем перевода предлагаются различные способы предотвращения семантических ошибок. Система ЭТАП в режиме диалога предлагает выбрать подходящий вариант для многозначного слова или омонима.

Варианты перевода от систем МП

СМП	Вариант машинного перевода	Семантические ошибки
Yandex	«I knew it! "he muttered in confusion, " I thought so! It's the worst of all! This is a kind of some foolishness, some пошлейшая* little thing**, the whole concept*** of may spoil. Yes, too conspicuous hat... Funny, therefore, noticeable... To my rags go will need a cap, at least the old pancake some, but not this monster****. None of these is not a place***** no one will notice, remember... the main thing, then remember, Academy of science and evidence*****...»	*<i>пошлейшая</i> (превосходная степень, здесь в значении 'тривиальный') **<i>little thing</i> (мелочь в значении 'пустяк' – <i>trifle</i>) ***<i>concept</i> (имеет значение 'авторский замысел', лучше <i>plan</i> в значении 'намерение') ****<i>monster</i> (относится к людям, животным, указывает на огромный размер, лучше <i>malformed thing</i>) ***** <i>a place</i> (за версту переводится как <i>a mile away</i>) ***** <i>Academy of science and evidence</i>(ан – союз, употребляется для противопоставления)
Prompt	"I and knew! — he muttered in confusion — I and thought! It is nastier than everything! Here the such any nonsense*, any poshleyshy** trifle, all plan can spoil! Yes, too perceptible hat... Ridiculous, therefore also the perceptible... To my tatter the peak-cap, at least an old pancake any, instead of this freak is by all means necessary. Anybody such doesn't carry***, far off**** will notice, will remember... the main thing, then will remember, An***** and a proof*****..."	*<i>nonsense</i> (в значении 'глупое поведение', лучше в значении 'глупый поступок' – <i>foolish / stupid / silly thing</i>) **<i>poshleyshy</i> (превосходная степень прилагательного <i>пошлый</i> в значении 'тривиальный') ***<i>carry</i> (в значении 'переносить', лучше <i>wear</i> в значении 'носить одежду') ****<i>far off</i> (нефразеологический вариант, лучше <i>a mile away / off</i>) *****<i>An</i> (ан – союз, употребляется для противопоставления) *****<i>proof</i> (официальный вариант, лучше <i>clue</i>)
ЭТАП	"I so also knew! -He muttered in embarrassment - I so also thought! Already in all it is worse! Such some foolishness, some most platitudinous trifle here are, the whole design* can spoil! Yes too visible a softy**... Funny therefore and visible... The peaked cap, at least old pancake some, and not this freak to my rags are certainly necessary. However also an evidence nobody doesn't wear, behind verst*** will be noticed, will memorize... the main thing, afterwards will memorize, such... "	*<i>design</i> (замысел в значении 'намерение' переводится как <i>plan</i>) **<i>softy</i> (в значении 'тряпка' о человеке, лучше <i>hat</i>) ***<i>behind verst</i> (пословный перевод, лучше <i>a mile away / off</i>)

Статистические системы способны самообучаться: чем больше в распоряжении имеется параллельных корпусов и чем точнее они соответствуют друг другу, тем лучше результат статистического машинного перевода, также пользователи сами могут пополнять базу знаний системы путем исправления ошибок в машинном переводе. Некоторые системы можно настраивать на определенные предметные области, что уменьшает вероятность появления семантических ошибок.

На сегодняшний момент производители систем МП разрабатывают гибридные системы на основе совмещения методов на примерах и правилах и статистических методов, беря в расчет сильные стороны этих технологий.

Литература

1. *Марчук Ю. Н.* Модели перевода: учеб. пособие для студ. учреждений высш. проф. образования. М.: Издательский центр «Академия», 2010.
2. *Баранов А. Н.* Введение в прикладную лингвистику: Учебное пособие. М.: Эдиториал УРСС, 2001.
3. *Harold Somer* Review Article: Example-based Machine Translation // Machine Translation. Kluwer Academic Publishers. Printed in the Netherlands, 2001. P. 113–157. Интернет-адрес: <http://kitt.cl.uzh.ch/clab/satzaehnlichkeit/tutorial/Unterlagen/Somers1999.pdf>.
4. Интернет-адрес: http://www.promt.ru/company/technology/machine_translation.
5. Интернет-адрес: <http://company.yandex.ru/technologies/translation>.

КОНЦЕПЦИЯ ПРОЕКТА ВЫСТАВОЧНОГО ПРОСТРАНСТВА «ПУЛЬС ЧЕРНОБЫЛЯ» (О ПОСЛЕДСТВИЯХ КРУПНЕЙШЕЙ ТЕХНОГЕННОЙ КАТАСТРОФЫ ДЛЯ БЕЛАРУСИ)

Н. В. Осадчий

Тема выставочного пространства «Пульс Чернобыля» связана с крупнейшей техногенной катастрофой в мировой истории и последствиями этой катастрофы для Беларуси. Мое обращение к данной теме и созданию проекта выставочного пространства, посвященного этой трагедии, было обусловлено осознанием глобальных проблем современного общества. Сравнительно недавно, примерно двести лет назад, произошла научно-техническая революция, но в этой прочной и крепкой на вид цепи начинают все чаще и чаще обнаруживаться ее слабые звенья, оставляющие глубокие шрамы на лице человеческого существования, что начинает пробуждать в нас объективное видение побочных явлений нашего комфортного проживания за счет технических достижений.

На создание проекта выставочного пространства «Пульс Чернобыля» повлияло также и то, что в самой Беларуси, которая особенно пострадала от этой техногенной катастрофы, до сих пор нет целостного визуаль-