

## РЕШЕНИЕ ЗАДАЧИ АВТОМАТИЧЕСКОГО ПЕРЕВОДА НАУЧНО-ТЕХНИЧЕСКИХ ТЕКСТОВ С АНГЛИЙСКОГО ЯЗЫКА НА БЕЛОРУССКИЙ

Ю. Б. Крапивин<sup>1</sup>, Д. В. Степанова<sup>2</sup><sup>1</sup> К. т. н., доцент кафедры интеллектуальных информационных технологий учреждения образования «Брестский государственный технический университет», Брест, Беларусь, e-mail: ubox@list.ru<sup>2</sup> К. ф. н., доцент, заместитель декана факультета английского языка учреждения образования «Минский государственный лингвистический университет», Минск, Беларусь, e-mail: daryastepanova79@gmail.com**Реферат**

Статья посвящена проблеме совершенствования качества машинного перевода научно-технических текстов. Рассматривается применение нейросетевого и статистического подходов с контролем по созданным автоматическим терминологическим словарям для разработки системы автоматического перевода английских научно-технических текстов по информационным технологиям на белорусский язык. Разработанная автоматизированная информационная система «Англо-белорусский словарь» обеспечивает автоматический лингвистический анализ входного текста и автоматический перевод текста выбранной предметной области с английского языка на белорусский с учетом информации, содержащейся в разработанных терминологических словарях лингвистической базы данных. Результаты исследования могут быть использованы при создании систем автоматического перевода терминологической лексики различных предметных областей, а также в разработках по совершенствованию существующих систем машинного перевода текстов.

**Ключевые слова:** машинный перевод, информационные технологии, терминологический словарь, нейросетевой подход, статистический подход, автоматизированная информационная система

## SOLUTION TO THE PROBLEM OF SCIENTIFIC AND TECHNICAL AUTOMATIC TRANSLATION FROM ENGLISH INTO BELARUSIAN

Yu. B. Krapivin, D. V. Stepanova

**Abstract**

The article deals with the problem of improving the quality of machine-based translation. The paper provides the neural network and statistical approaches with control over the created automatic specialized dictionaries for the development of a system for the automatic translation of English scientific and technical texts on information technologies into Belarusian. The article considers the main aspects and stages of linguistic database and algorithmic model developing for the given purpose and analyses the results and prospects of the developed automated information system "English-Belarusian Dictionary".

**Keywords:** machine translation, information technologies, specialized dictionaries, neural network approach, statistical approach, automated information system

**Введение**

Идея машинного перевода разными темпами разрабатывается с пятидесятих годов двадцатого столетия. Машинный или автоматический перевод можно определить как выполняемое компьютером действие по преобразованию текста на одном естественном языке в текст на другом естественном языке при сохранении содержания, а также результат такого действия [1, с. 80].

В последнее время в мировой практике особое внимание уделяется совершенствованию качества оперативного машинного перевода научно-технических текстов. Актуальность данного направления исследования обусловлена рядом обстоятельств. Во-первых, это существующая в системе информационного обслуживания необходимость оперативной обработки постоянно возрастающих объемов научно-технической информации, поступающей из иностранных источников в виде специальной литературы, аннотаций, рефератов обзоров, технических деловых документов. И во-вторых, это современная проблематика становления, развития и взаимодействия терминологических систем различных областей науки и техники, а также взаимосвязь и взаимопроникновение общеупотребительной лексики и терминологии в ходе межкультурной коммуникации.

Наряду с такими важными для автоматического перевода параметрами, как емкость и быстродействие компьютеров, оптимальность работы с текстовыми данными, главными в проблеме машинного перевода остаются проблемы формализации лингвистических данных от простой морфологической структуры слова до сложной семантической организации целого текста.

В представленной работе рассматривается возможность использования нейросетевого и статистического подходов к автоматическому переводу научно-технических текстов по информационным технологиям с английского языка на белорусский. Применение комбинированного подхода к решаемой задаче по совершенствованию

машинного перевода предполагает обязательный контроль по разработанному автоматическому словарю терминологической лексики по информационным технологиям.

Выбор предметной области «Информационные технологии» для проведения исследования не случаен и объясняется тем, что средства компьютерной техники, программное обеспечение, глобальная сеть Интернет характеризуются стремительными темпами развития и увеличением масштаба применения, что сопровождается разработкой и внедрением новых технологий, требующих номинации. В результате бумажные словари не успевают отразить все изменения в составе терминологии данной сферы, что вызывает определенную сложность при переводе. Кроме того, обеспечение массового доступа к компьютерным технологиям приводит к расширению сферы употребления терминологической лексики выбранной предметной области.

Решение задачи автоматического перевода научно-технических текстов по информационным технологиям с английского языка на белорусский соответствует приоритетному фундаментальному и прикладному научному направлению исследований в Республике Беларусь.

**Основная часть Реализация автоматизированной информационной системы (АИС) «Англо-белорусский словарь»**

К настоящему времени существует достаточно много методов, алгоритмов и практических систем машинного перевода текстов различных естественных языков. Так, например, в работах [2, 3] достаточно полно проведен анализ методов машинного перевода. В частности, в [3, с. 31] приводятся три основные стратегии разработки систем машинного перевода:

- 1) стратегия «прямого» перевода (direct translation systems): в этом случае алгоритмы перевода рассчитаны на конкретную пару языков и сравнительно неглубокий анализ входного текста без фиксирования его лингвистической структуры и четкого разделения

- процессов анализа и синтеза. Процесс перевода требует минимума преобразований: входной текст постепенно превращается в текст на выходном языке путём замены всех его элементов, найденных в словаре, на их переводные эквиваленты. Такой перевод ещё называют пословным или пословно-оборотным;
- 2) «трансферная» стратегия (transfer systems): в процессе анализа входного текста определённым образом формируется его внутреннее представление, затем осуществляется преобразование этого представления как на лексическом, так и на структурном уровнях, в соответствующие структуры выходного языка и далее выполняется лексический синтез. Процессы анализа и синтеза в таких системах являются независимыми. Анализ ведётся в категориях входного языка, а синтез – в категориях выходного языка; связь того и другого этапов обеспечивается введением особого этапа межязыковых операций (собственно перевода – трансфера);
  - 3) перевод с использованием «интерлингвы» – языка-посредника (interlingua systems): здесь используются самые разнообразные методы анализа естественного языка с тем, чтобы отобразить содержание входного текста в некоторое универсальное представление, не зависящее в идеале ни от входного, ни от выходного языков; то есть сам процесс перевода фактически включает только две последовательные «чистые» фазы – анализ и синтез.

При этом указанные стратегии реализуются посредством следующих основных методов: статистических [4–7], одной из разновидностей которых являются методы, основанные на примерах (example-based – EB-методы) [8–16], и лингвистических, к примеру, методы, основанные на правилах – rule-based – RB-методы [17–19]. Первые – сводятся к использованию параллельных корпусов текстов, составляющих основу для вычисления соответствия друг другу лексических единиц различных языков и их статистических характеристик, а также построения некоторой модели перевода. Вторые, в свою очередь, – к построению модели перевода в соответствии с набором лингвистических правил, задающих нужную глубину анализа текста, а также возможную трансформацию грамматической структуры входного текста в эквивалентные ей структуры выходного текста.

Также приводится характеристика наиболее известных промышленных систем машинного перевода, представляющих собой практическую реализацию описанных методов. К ним относятся:

- 1) система Promt [20], которая является системой машинного перевода трансферного типа и позволяет получать перевод текстов между несколькими европейскими языками (всего 24 языковых направления). В основе работы системы лежит комбинация EB- и лингвистических методов автоматического перевода. Её базовые словари содержат от 40 до 200 тысяч слов. Также предусмотрены наборы специализированных словарей, ориентированные на определённые предметные области (например, для англо-русского перевода используется более 50 тематических словарей). Скорость перевода составляет более 1 миллиона слов в час;
- 2) система SysTran [21], осуществляющая перевод между несколькими европейскими языками. В её основе также используется

- комбинация EB- и лингвистических методов машинного перевода. Разрешение проблемы многозначности лексических единиц производится на основе анализа их контекста. Скорость перевода составляет более 3 миллионов слов в час;
- 3) система Retrans [22], которая является системой фразеологического типа и основана на выполнении поверхностного семантико-синтаксического анализа текста с целью выделения фраз (слов и словосочетаний), последующим их переводом на основе словарей и дальнейшим синтезом выходного текста с применением синтаксических и морфологических алгоритмов. Общий объём базового словаря составляет около 4 миллионов словарных статей. Также доступны дополнительные тематические словари объёмом около 400 тысяч словарных статей. Скорость перевода составляет порядка 432 тысячи слов в час;
  - 4) система Белазар [23] представляет собой систему русско-белорусского перевода, базовый словарь которой составляет около 78 тысяч канонических словоформ для каждого из языков;
  - 5) системы on-line перевода Yandex.Переводчик [24] и Google.Translate [25] также построены на применении статистических методов и обеспечивают возможность автоматического перевода как отдельных фрагментов текста, так и целых веб-страниц, представленных на 95 и 103 языках, соответственно.

Проведенный анализ систем машинного перевода дает основание говорить о значительных результатах, достигнутых специалистами в области компьютерной лингвистики и автоматической обработки текста. К примеру, в системах машинного перевода наблюдается тенденция увеличения числа рабочих языков, создания дополнительных словарей и средств для перевода текстов непосредственно в сети Интернет. Во многих системах реализован подход к языку, как к единому целому, обеспечивающий учет описания морфологических, синтаксических и семантических связей. Однако, несмотря на все положительные стороны и возможности существующих систем машинного перевода, некоторые лингвистические проблемы остаются нерешенными. Например, находящиеся в свободном доступе в сети Интернет системы машинного перевода достаточно «чувствительны» к узкоспециализированной лексике и не всегда обеспечивают качественный перевод, в связи с чем представляется обоснованным применение автоматических терминологических словарей как неотъемлемого компонента системы машинного перевода в целях совершенствования результатов ее работы.

Разработанная информационная система «Англо-белорусский словарь» обеспечивает автоматизацию процесса ведения словарей терминологической лексики по информационным технологиям, а также пословно-оборотного перевода английских текстов по информационным технологиям на белорусский язык в человеко-машинном режиме взаимодействия. Система состоит из двух подсистем (рисунок 1):

- подсистема ведения словарей терминологической лексики по информационным технологиям;
- подсистема пословно-оборотного перевода английских текстов по информационным технологиям на белорусский язык.

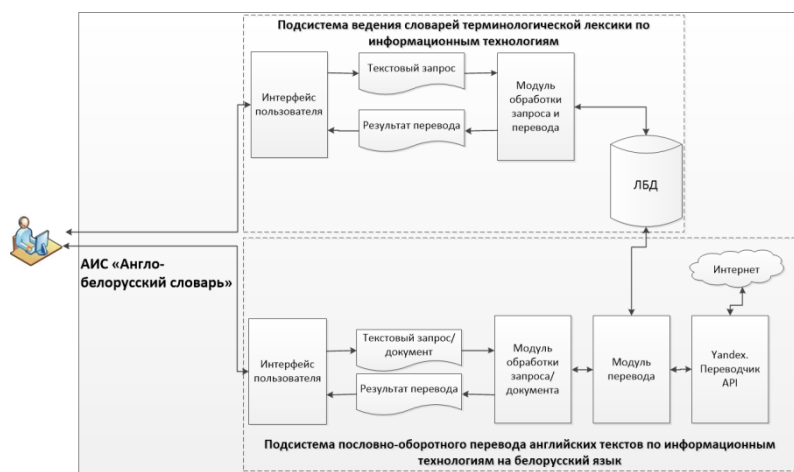


Рисунок 1 – Структурно-функциональная схема АИС «Англо-белорусский словарь»

Входной информацией для системы являются подлежащие переводу на белорусский язык английские тексты по информационным технологиям. При этом перевод таких текстов может осуществляться как в человеко-машинном диалоговом режиме взаимодействия, так и в автоматическом режиме. Основу работы человеко-машинного диалогового режима обеспечивает подсистема ведения словарей терминологической лексики по информационным технологиям. В данном случае с целью поиска термина и его варианта перевода специалист обращается через Интерфейс пользователя посредством модуля обработки запроса и перевода к созданному словарю из лингвистической базы данных (ЛБД). Автоматический режим опирается на подсистему пословно-оборотного перевода английских текстов по информационным технологиям на белорусский язык и предполагает выполняемый компьютером в модуле обработки запроса/документа анализ поступающего через Интерфейс пользователя текста или его фрагмента, например в несколько предложений, с целью последующего перевода содержащейся в нем терминологической (с помощью терминологических словарей из ЛБД) и общеупотребительной (функциональность Yandex.Переводчик API) лексики. Результат перевода возвращается пользователю для дальнейшего анализа и принятия решения.

Функциональность подсистемы ведения словарей терминологической лексики по информационным технологиям обеспечивает взаимодействие с ЛБД с целью выполнения вызываемых из модуля обработки запроса и перевода операций по навигации, добавлению, удалению, изменению, поиску по заданному условию, фильтрации, автоматической замене, сохранению и печати записей из разработанных англо-белорусских двуязычных словарей по информационным технологиям (фрагмент словаря приведен на рисунке 2):

- терминологического словаря;
- терминологического словаря словосочетаний;
- словаря аббревиатур и сокращений.

Функциональность подсистемы пословно-оборотного перевода английских текстов по информационным технологиям на белорусский язык базируется на применении Yandex.Переводчик API, реализующего гибридную модель машинного перевода, что, в свою очередь, предполагает комбинирование нейросетевого и статистического подходов с контролем по словарям терминологической лексики.

Так, согласно [26], статистический подход основан на моделях языка и перевода. Для создания модели перевода выполняется сравнение сотен тысяч параллельных текстов, одинаковых по смыслу, но написанных на разных языках. Сравнивая их, система учится находить соответствия и, например, запоминает, что слова «ability» и «здольнасць» являются вероятными переводами друг друга. Результат сравнения записывается в матрицу выравнивания слов.

Матрица помогает определить, какие пары фраз в паре предложений могут служить переводами друг для друга. Для создания модели языка выполняется анализ текстов на одном языке и составляются списки всех употребляемых слов и фраз. Каждому слову и фразе соответствует свой числовой идентификатор, определяющий их «статистическую популярность в языке» (частоту использования в проанализированных текстах). В процессе перевода каждое исходное предложение разделяется на слова и фразы, которые переводятся независимо друг от друга. Для каждой части предложения подбирается потенциальный перевод из матрицы. Затем выполняется повторное построение нескольких вариантов предложения и выбор наилучшего варианта (с учетом имеющихся сведений о частоте употребления в проанализированных текстах) с точки зрения оптимальной сочетаемости слов естественного языка. Достоинством статистического подхода является способность запоминать и переводить короткие фразы и слова с невысоким показателем частотности употребления, а в качестве недостатка можно отметить отсутствие взаимосвязи между фразами, так как при переводе не учитывается контекст.

В рамках нейросетевого подхода [26] нейронная сеть также анализирует массив параллельных текстов, учится находить в них закономерности и составляет списки всех употребляемых слов и фраз. Однако вместо простых идентификаторов, используемых при статистическом подходе, в нейросетевом – используется векторное представление слов или word embedding. Вектор состоит из чисел, характеризующих слово по лексическим и семантическим признакам. Во время перевода каждое исходное предложение не разбивается на слова и фразы, а переводится полностью. Каждому слову в предложении сопоставляется вектор длиной в несколько сотен чисел, тем самым формируя векторное пространство. В этом векторном пространстве нейронная сеть определяет семантику слов и взаимосвязь между ними, даже если слова находятся в разных частях исходного предложения. Это, например, позволяет установить, что слова «abort» и «abandon» часто появляются в сходных контекстах. Оба этих слова могут находиться в контексте нового слова «program». И если в обучающих данных со словом «program» встречается лишь одно из них (например, «abort»), то выбор будет сделан в пользу последнего. Достоинством нейросетевого подхода является способность учитывать взаимосвязь между словами, что позволяет добиться более связанного перевода, тогда как недостатком является отсутствие достаточной информации по словам, которые редко встречались в текстах обучающей выборки и для которых система еще не смогла построить приемлемое векторное представление, например термины узкоспециализированной лексики.

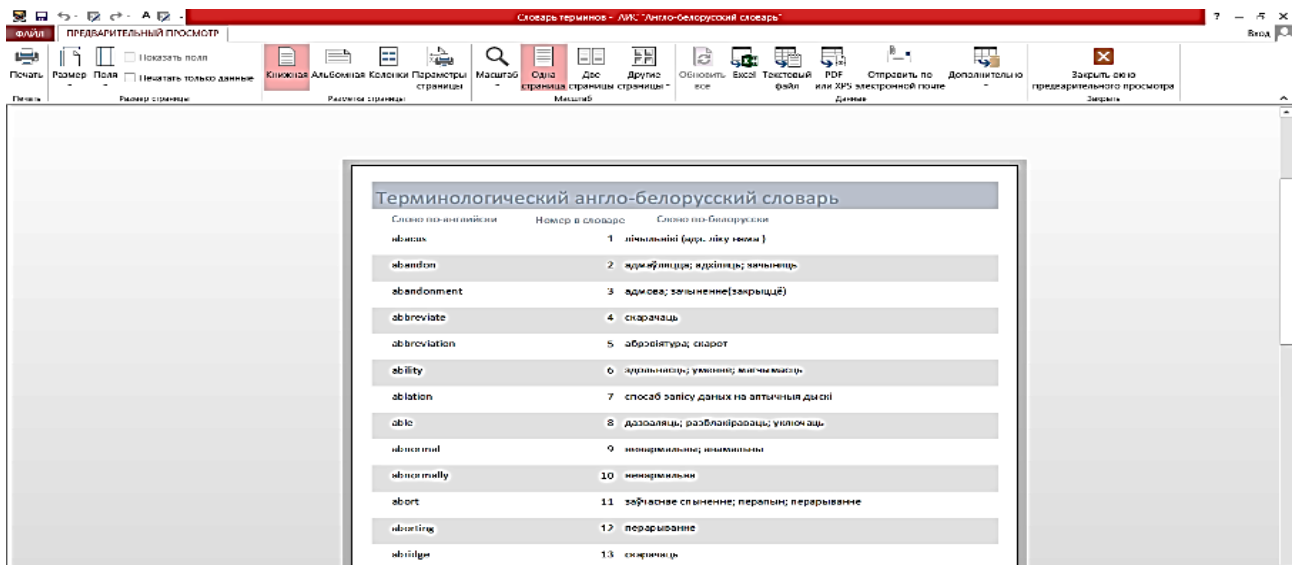


Рисунок 2 – Фрагмент англо-белорусского двуязычного словаря по информационным технологиям

Комбинирование рассмотренных выше подходов обеспечивается с помощью алгоритма анализа результатов, получаемых от их применения в отдельности, основанном на методе обучения CatBoost [27]. При вычислении оценки учитываются десятки факторов – от длины предложения (короткие фразы и слова с невысоким показателем частотности употребления лучше переводятся при использовании статистического подхода) до анализа синтаксической структуры. Алгоритм позволяет оценить оба перевода по всем факторам, выбирая лучший и возвращая его в качестве результата.

Реализованная в рамках подсистемы пословно-оборотного перевода английских текстов по информационным технологиям на белорусский язык (модуль обработки запроса/документа, модуль перевода) возможность первоочередного поиска и перевода терминов входного текста на основании терминологических словарей из лингвистической базы данных также позволяет обеспечить:

- автоматический лингвистический анализ входного текста в части сбора сведений о частоте встречаемости словоформ из словарей в тексте, их лемматизации, определения принадлежности к части речи, визуализации дерева грамматического разбора заданного предложения и сохранения собранной информации в соответствующие файлы на жестком диске;
- автоматический перевод текста выбранной предметной области с английского языка на белорусский с учетом информации,

содержащейся в разработанных терминологических словарях лингвистической базы данных созданной системы.

Так, например, ниже представлен фрагмент результатов лингвистического анализа входного текста.

Таблица 1 – Фрагмент результатов автоматического лингвистического анализа входного текста на английском языке

Словоформа	Лемма	Тег части речи
ability	Ability	NOUN
arithmetic	Arithmetic	ADJ
....	...	...
computers	Computer	NOUN
...	...	...
follow	Follow	VERB
...	...	...

На рисунке 3 приведено дерево грамматического разбора одного из предложений входного текста, обработанного в подсистеме пословно-оборотного перевода английских текстов по информационным технологиям на белорусский язык и размеченного с помощью лексико-грамматических кодов набора Penn Treebank [28].

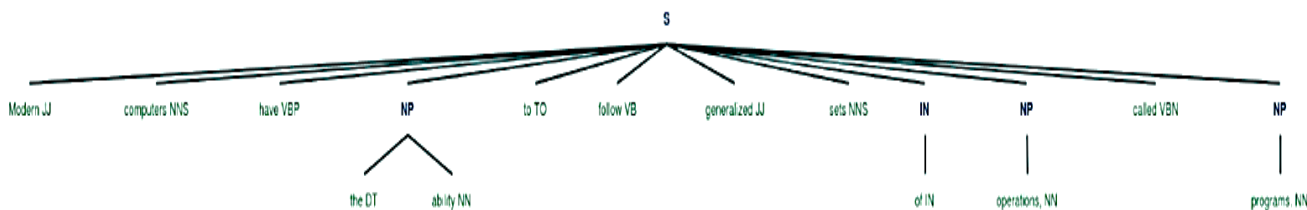


Рисунок 3 – Дерево грамматического разбора предложения «Modern computers have the ability to follow generalized sets of operations, called programs.»

**Заключение**

Таким образом, разработанная автоматизированная информационная система «Англо-белорусский словарь» позволяет усовершенствовать существующие алгоритмы автоматического перевода научно-технических текстов, а также дополнять и обновлять созданную лингвистическую базу данных.

Предложенная гибридная модель перевода, основанная на нейросетевом и статистическом методах с контролем по созданным терминологическим словарям, доказала высокую эффективность функциональности и может быть использована для формализации процесса перевода терминологической лексики других предметных областей, а также послужить основой для решения лингвистических проблем автоматического перевода.

Практическая значимость проведенного исследования заключается в возможности применения разработанной системы как в качестве самостоятельного программного обеспечения для перевода терминологической лексики предметной области «Информационные технологии», так и в качестве подпрограммы к уже существующим системам машинного перевода. Отобранные для автоматического словаря английские термины, терминологические словосочетания и аббревиатуры, а также их варианты перевода на белорусский язык могут быть использованы при создании переводных белорусско-русских, белорусско-английских и англо-белорусских словарей, при написании учебников и учебных пособий по переводу, а также в преподавании научно-технического перевода.

**Список цитированных источников**

1. Марчук, Ю. Н. Научно-технический перевод / Ю. Н. Марчук. – М. : Наука, 1987. – 144 с.
2. Крапивин, Ю. Б. Методы и алгоритмы автоматического распознавания воспроизведенных фрагментов текстовых документов : дис. канд. тех. наук : 05.13.17 / Ю. Б. Крапивин. – Минск, 2019. – 129 л.

3. Воронков, Н. В. Методы, алгоритмы и модели систем автоматического реферирования текстовых документов : дис. канд. тех. наук : 05.13.17 / Н. В. Воронков – Минск, 2007. – 165 л.
4. A Statistical Approach to Machine Translation / P.F. Brown [et al.] // Computational Linguistics. – 1990. – Vol. 16, № 2. – P. 79–85.
5. Statistical Methods for Machine Translation / S. Vogel [et al.] // Verbmobil: Foundations of Speech-to-Speech Translation / W. Wahlster; ed.: W. Wahlster. – Springer Verlag: Berlin, 2000. – P. 377–393.
6. Melamed, I. A Word-to-Word Model of Translational Equivalence / I. Melamed // Association for Computational Linguistics: proceedings of the 35th Annual Meeting, Madrid, 7-12 July 1997 / ACL Press, 1997. – P. 490–497.
7. Translation word learning scheme for machine translation: US Patent №5826220, G06F 17/28 (20060101); G06F 017/28 / K. Takeda, Y. Saito, H. Hirakawa; Assignee: Kabushiki Kaisha Toshiba; filed 02.10.1995; pub. 20.10.1998. – 1998. – P. 20.
8. Schaler, R. Beyond Translation Memories / R. Schaler // Workshop on Example-Based Machine Translation [Electronic resource]. – 2014. – Mode of access: <http://www.eamt.org/events/summitVIII/papers/schaeler.pdf>. – Date of access: 8.02.2014.
9. Nagao, M. A framework of a mechanical translation between Japanese and English by analogy principle / M. Nagao // Artificial and human intelligence: proceedings of the International NATO symposium, Amsterdam, 1984 / NATO. – Elsevier North-Holland, Inc., 1984. – P. 173–180.
10. Turcato, D. What is Example-Based Machine Translation? / D. Turcato, F. Popowich // Recent Advances of EBMT / M. Carl, A. Way; ed.: M. Carl [et al.]. – Kluwer Academic Publishers, 2003. – P. 59–82.
11. Somers, H. Review Article: Example-based Machine Translation / H. Somers // Machine Translation. – 1999. – Vol. 14, № 2. – P. 113–157.
12. Deriving Transfer Rules from Dominance-Preserving Alignments / A. Meyers [et al.] // Association for Computational Linguistics: proceedings of the 36th Annual Meeting, Montreal, 10-14 August 1998 / ACL. – Morgan Kaufmann Publishers, 1998. – P. 843–847.

13. Matsumoto, Y. Acquisition of Translation Rules from Parallel Corpora / Y. Matsumoto, M. Kitamura // Recent Advances in Natural Language Processing: Selected Papers from RANLP'95, Tzigras Chark, 14–16 September 1995 / R. Mitkovand, N. Nicolov; ed.: R. Mitkov [et al.]. – John Benjamins Publishing Co., 1997. – P. 405–416.
14. Carroll, J. J. Repetitions Processing Using a Metric Space and the Angle of Similarity / J. J. Carroll. – Report No. 90/3, Centre for Computational Linguistics, UMIST, Manchester, England, 1990. – 27 p.
15. Brown, R. D. Example-Based Machine Translation in the Pangloss System / R. D. Brown // Computational Linguistics: proceedings of the 16th Conf., Copenhagen, 1996 / ACL. – ACL, 1996. – P. 169–174.
16. Brockett, C. English-Japanese Example-Based Machine Translation Using Abstract Linguistic Representation / C. Brockett [et al.] // Machine translation in Asia: proceedings of the Workshop, Taipei, 24 August – 1 September, 2002 / Academia Sinica, ACLCLP, National Tsing Hua University. – Morgan Kaufmann Publishers, 2002. – P. 19–26.
17. Kaji, H. An Efficient Method for Rule-Based Machine Translation / H. Kaji // Computational Linguistics: proceedings of the 12th Conf., Budapest, 22–27 August 1988 / ACL. – ACL, 1988. – P. 824–829.
18. Jassem, K. POLENG - Adjusting a Rule-Based Polish-English Machine Translation System by Means of Corpus Analysis / K. Jassem, F. Gralinski, G. Krynicki // Harvesting existing resources: proceedings of the 5th EAMT Workshop, Ljubljana, 11–12 May 2000 / European Association for Machine Translation. – MIT Press Cambridge, 2000. – P. 47–50.
19. Tamas, G. MetaMorpho TM: A Rule-Based Translation Corpus / G. Tamas, H. Gabor, K. Balazs // Language Resources and Evaluation: proceedings of the 4th International Conf., Lisbon, 26–28 May 2004 / ELRA. – ELRA, 2004. – P. 339–342.
20. PROMT Translator / PROMT Translation Software and Dictionaries [Electronic resource]. – 2020. – Mode of access : <http://www.promt.com>. – Date of access : 11.11.2020.
21. SYSTRAN – Translation Technologies: Online translation, translation software and tools / SYSTRAN [Electronic resource]. – 2020. – Mode of access : <http://www.systransoft.com>. – Date of access : 12.12.2020.
22. RETRANS / Ретранс Текнолоджиз [Электронный ресурс]. – 2020. – Режим доступа : <http://www.retrans.ru/>. – Дата доступа : 9.12.2020.
23. Белазар – беларуска-рускі перакладнік / Белазар [Электронный ресурс]. – 2020. – Режим доступа : <http://belazar.info>. – Дата доступа : 19.12.2020.
24. Yandex.Переводчик [Электронный ресурс]. – 2021. – Режим доступа : <https://translate.yandex.by/help>. – Дата доступа : 7.01.2021.
25. Google.Translate [Electronic resource]. – 2021. – Mode of access : <https://translate.google.com/intl/en/about/>. – Date of access : 7.01.2021.
26. О машинном переводе. – Руководство разработчика [Электронный ресурс]. – Режим доступа : <https://yandex.ru/dev/translate/doc/dg/concepts/how-works-machine-translation.html>. – Дата доступа : 7.01.2021.
27. Overview of CatBoost – CatBoost. Documentation [Electronic resource]. – 2021. – Mode of access : <https://catboost.ai/docs/>. – Date of access : 7.01.2021.
28. Taylor, A., The Penn Treebank: An Overview / A. Taylor, M. Marcus, B. Santorini // Treebanks: Building and Using Parsed Corpora / ed.: Abeillé, A. – Springer Netherlands, 2003. – P. 5–22.
8. Schaler, R. Beyond Translation Memories / R. Schaler // Workshop on Example-Based Machine Translation [Electronic resource]. – 2014. – Mode of access: <http://www.eamt.org/events/summitVIII/papers/schaeler.pdf>. – Date of access: 8.02.2014.
9. Nagao, M. A framework of a mechanical translation between Japanese and English by analogy principle / M. Nagao // Artificial and human intelligence: proceedings of the International NATO symposium, Amsterdam, 1984 / NATO. – Elsevier North-Holland, Inc., 1984. – P. 173–180.
10. Turcato, F. Popowich // Recent Advances of EBMT / M. Carl, A. Way; ed.: M. Carl [et al.]. – Kluwer Academic Publishers, 2003. – P. 59–82.
11. Somers, H. Review Article: Example-based Machine Translation / H. Somers // Machine Translation. – 1999. – Vol. 14, № 2. – P. 113–157.
12. Deriving Transfer Rules from Dominance-Preserving Alignments / A. Meyers [et al.] // Association for Computational Linguistics: proceedings of the 36th Annual Meeting, Montreal, 10–14 August 1998 / ACL. – Morgan Kaufmann Publishers, 1998. – P. 843–847.
13. Matsumoto, Y. Acquisition of Translation Rules from Parallel Corpora / Y. Matsumoto, M. Kitamura // Recent Advances in Natural Language Processing: Selected Papers from RANLP'95, Tzigras Chark, 14–16 September 1995 / R. Mitkovand, N. Nicolov; ed.: R. Mitkov [et al.]. – John Benjamins Publishing Co., 1997. – P. 405–416.
14. Carroll, J. J. Repetitions Processing Using a Metric Space and the Angle of Similarity / J. J. Carroll. – Report No. 90/3, Centre for Computational Linguistics, UMIST, Manchester, England, 1990. – 27 p.
15. Brown, R. D. Example-Based Machine Translation in the Pangloss System / R. D. Brown // Computational Linguistics: proceedings of the 16th Conf., Copenhagen, 1996 / ACL. – ACL, 1996. – P. 169–174.
16. Brockett, C. English-Japanese Example-Based Machine Translation Using Abstract Linguistic Representation / C. Brockett [et al.] // Machine translation in Asia: proceedings of the Workshop, Taipei, 24 August – 1 September, 2002 / Academia Sinica, ACLCLP, National Tsing Hua University. – Morgan Kaufmann Publishers, 2002. – P. 19–26.
17. Kaji, H. An Efficient Method for Rule-Based Machine Translation / H. Kaji // Computational Linguistics: proceedings of the 12th Conf., Budapest, 22–27 August 1988 / ACL. – ACL, 1988. – P. 824–829.
18. Jassem, K. POLENG - Adjusting a Rule-Based Polish-English Machine Translation System by Means of Corpus Analysis / K. Jassem, F. Gralinski, G. Krynicki // Harvesting existing resources: proceedings of the 5th EAMT Workshop, Ljubljana, 11–12 May 2000 / European Association for Machine Translation. – MIT Press Cambridge, 2000. – P. 47–50.
19. Tamas, G. MetaMorpho TM: A Rule-Based Translation Corpus / G. Tamas, H. Gabor, K. Balazs // Language Resources and Evaluation: proceedings of the 4th International Conf., Lisbon, 26–28 May 2004 / ELRA. – ELRA, 2004. – P. 339–342.
20. PROMT Translator / PROMT Translation Software and Dictionaries [Electronic resource]. – 2020. – Mode of access : <http://www.promt.com>. – Date of access : 11.11.2020.
21. SYSTRAN – Translation Technologies: Online translation, translation software and tools / SYSTRAN [Electronic resource]. – 2020. – Mode of access : <http://www.systransoft.com>. – Date of access : 12.12.2020.
22. RETRANS / Retrans Teknologizhiz [Elektronnyj resurs]. – 2020. – Rezhim dostupa : <http://www.retrans.ru/>. – Data dostupa : 9.12.2020.
23. Belazar – belarуска-рускі перакладнік / Belazar [Elektronnyj resurs]. – 2020. – Re-zhim dostupa : <http://belazar.info>. – Data dostupa : 19.12.2020.
24. Yandex.Perevodchik [Elektronnyj resurs]. – 2021. – Rezhim dostupa : <https://translate.yandex.by/help>. – Data dostupa : 7.01.2021.
25. Google.Translate [Electronic resource]. – 2021. – Mode of access : <https://translate.google.com/intl/en/about/>. – Date of access : 7.01.2021.
26. О машинном переводе. – Руководство разработчика [Elektronnyj resurs]. – Rezhim dostupa : <https://yandex.ru/dev/translate/doc/dg/concepts/how-works-machine-translation.html>. – Data dostupa : 7.01.2021.
27. Overview of CatBoost – CatBoost. Documentation [Electronic resource]. – 2021. – Mode of access : <https://catboost.ai/docs/>. – Date of access : 7.01.2021.
28. Taylor, A., The Penn Treebank: An Overview / A. Taylor, M. Marcus, B. Santorini // Treebanks: Building and Using Parsed Corpora / ed.: Abeillé, A. – Springer Netherlands, 2003. – P. 5–22.

#### References

1. Marchuk, Yu. N. Nauchno-tehnicheskij perevod / Yu. N. Marchuk. – M.: Nauka, 1987. – 144 s.
2. Krapivin, Yu. B. Metody i algoritmy avtomaticheskogo raspoznavaniya vosproizvedennykh fragmentov tekstovykh dokumentov : dis. kand. tekhn. nauk : 05.13.17 / Yu.B. Krapivin. – Minsk, 2019. – 129 l.
3. Voronkov, N. V. Metody, algoritmy i modeli sistem avtomaticheskogo referirovaniya tekstovykh dokumentov : dis. kand. tekhn. nauk : 05.13.17 / N. V. Voronkov – Minsk, 2007. – 165 l.
4. A Statistical Approach to Machine Translation / P.F. Brown [et al.] // Computational Linguistics. – 1990. – Vol. 16, № 2. – P. 79–85.
5. Statistical Methods for Machine Translation / S. Vogel [et al.] // Verbmobil: Foundations of Speech-to-Speech Translation / W. Wahlster; ed.: W. Wahlster. – Springer Verlag: Berlin, 2000. – P. 377–393.
6. Melamed, I. A Word-to-Word Model of Translational Equivalence / I. Melamed // Association for Computational Linguistics: proceedings of the 35th Annual Meeting, Madrid, 7–12 July 1997 / ACL. – ACL Press, 1997. – P. 490–497.
7. Translation word learning scheme for machine translation: US Patent №5826220, G06F 17/28 (20060101); G06F 017/28 / K. Takeda, Y. Saito, H. Hirakawa; Assignee: Kabushiki Kaisha Toshiba; filed 02.10.1995; pub. 20.10.1998. – 1998. – P. 20.

Материал поступил в редакцию 25.02.2021