4. Вопросы теории и практики перевода: английский язык

**ОСОБЕННОСТИ РАЗРАБОТКИ УЗБЕКСКО-АНГЛИЙСКОГО ДВУЯЗЫЧНОГО ПРОГРАМНОГО КОМПЛЕКСА НА ОСНОВЕ МОДЕЛИРОВАНИЯ ГРАММАТИЧЕСКИХ КАТЕГОРИЙ ГЛАГОЛОВ И МОРФОЛОГИЧЕСКОГО АНАЛИЗА ПРИ МАШИННОМ ПЕРЕВОДЕ.**

**Расулова Умида**

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ**

**РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФГАОУ ВО «СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

**urasulova01@icloud.com**

**Аннотация.**

В данной статье показан путь разработки морфемного моделирования грамматических категорий глаголов узбекского языка для машинного перевода.

Также в статье почеркнуто именно на какие аспекты необходимо обратить внимание в момент их выражения на базе данных.

Указываются типы морфологического анализа грамматических категорий глагола, также обозначены общие парадигмы и различия в языке оригинала и перевода. Отличаются аналитические особенности и принципы формирования глаголов. Эти особенности глаголов, как и других частей речи определяют необходимость проведения специальных лингвистических исследований по сложным, вспомогательным глаголам, по всем видам словосочетаний, для введения модулей в программную базу, изучаемую автором для создания совершенного машинного перевода.

Так же в моделировании грамматических категорий глагола указаны их структура, способы соединений, состав, дефиниция комбинаций аффиксов глагола. Глаголы узбекского языка отличаются от глаголов английского языка их аналитическим характером и особой формой образования, т. е. сочетанием глаголов в форме совместных, вспомогательных; классификация видов требует крупного лингвистического анализа в компьютерной морфологии. Обогащение словарного запаса в узбекском языке и развитие науки и техники, рождающей неологизмы и вводящие заимствованные слова так же требуют отдельных исследований и анализа изучения.

Особое внимание уделяется построению предложений при переводе с английского на узбекский и обратно. Данная тема вполне может служить основой для новых исследований и научных работ, так требует более глубокого изучения внутренних языковых способностей узбекской словесной семьи и их лексико-семантических моделей.

**Ключевые слова:** разработка комплекса, узбекский язык, английский язык, двуязычный, программа, основа, модель, категория глаголов, морфологический анализ, машинный перевод, компьютерная лингвистика, автоматическая морфология, грамматическая категория, аналитические глаголы, словосочетание, морфологический анализ, моделирование, база данных.

**Введение.**

На сегодняшний день невозможно представить ни одну сферу деятельности без ИКТ. Информационные технологии играют большую роль в развитии науки и оказывают большое влияние на оптимизацию инфраструктуры в сфере накопления знаний, в основном на Интернет Технологии как во всём мире, так и в странах СНГ. Инфраструктура Узбекистана тоже претерпела большие изменения с обретением независимости. Наиболее сильно это отразилось в сфере образования. Широкое распространение вычислительных технологий и сети Интернет позволило решать серьезные научные проблемы. Появились новые научные направления, где используются компьютерные технологии.

Не мало примеров этого мы видим сегодня. Это и медицина, и педагогика, банки, сотрудники которых широко используют информационные технологии как в повседневной жизни, так и при научных докладах и исследованиях, проектах и тренингах.

Таким образом с 2000-х годов были оснащены и оборудованы лаборатории, ИРЦ ВОУ, программные классы в школах и средне-специальных образовательных учреждения. Одной из основных идеологий узбекской компьютерной лингвистики является максимально правильный машинный перевод, основанный на математическом моделировании и содержащий в себе все особенности английского и узбекского языков. Ведь именно перевод является тем самым ключом к прогрессу и быстрому обмену опытом и инновациями.

Необходимо отметить, что ряд учёных нашей республики проводили первые эксперименты и лабораторные исследования: PhD М.Хакимов проводил множество исследований по математическому моделированию в области многоязычного машинного перевода. Под руководством М.Хакимова созданы следующие справочники и учебные пособия: «Компьютерная лингвистика» (А. Пулатов, 2011), «Основы компьютерной лингвистики» (А.Рахимов, 2011), «Лингвистические основы машинного перевода» (Н.Абдурахмонова, 2012) и др. Большинство из этих работ имели теоретическую форму и содержание, и не было никаких программных реализаций, основанных на реальных лингвистических базах данных. Не проводилась совместная работа лингвистов с программистами. В настоящее время компьютерная лингвистика как наука преподается в нескольких государственных университетах Узбекистана: в

Ташкенте, Андижане, Намангане, Фергане, Хорезме, Самарканде, Бухаре.

Благодаря Указам Первого Президента и чётко обозначенным стратегиям ныне действующего Президента Ш.Мирзиёева поставлены акценты на следующие существенные вопросы: «... обеспечение соответствующего места нашего родного языка в мировой информационной сети Интернет, его компьютерное обеспечение и наличие научно-методических пособий, связанных с машинным переводом и электронными словарями, подготовка рекомендаций для широкого применения результатов на практике».

Узбекский язык – это язык великих философов, учёных, поэтов и писателей, признанных на мировой арене, таких как Улугбек, Ибн Сина, Алишер Навои, Аль Хорезмий и т.п.

Узбекский язык относится к тюркским языкам, который имеющим древнюю историю и постоянно меняющим своё состояние по различным причинам. Его отличительные особенности от других языков мы можем видеть на каждом из языковых уровней.

Например, сохраненная сингармония гласных на турецком языке в таких словах, как üzüm, velâyet. Однако в узбекском языке больше заимствований, чем в турецком.

Например, insulin, management, budget, test из английского, стол, поезд, бухгалтер из русского, agronomiya, allergiya, nargis из греческого, vazir, maktab, maorif из арабского.

С другой стороны, одной из важных задач компьютерной лингвистики в Узбекистане является создание грамматических анализаторов для узбекского языка, соответствующих латинской письменности. В соседних странах таких как Киргизия, Казахстан «…разработаны начальная версия системы морфологического анализатора в сфере Embercadero RAD Studio, с учетом морфологии. Она конечно же требует дальнейшего развития и ведутся научные разработки в данном направлении.»1 Точно так же в Узбекистане ряд молодых учёных стремится разработать и внедрить собственный узбекско- английский двуязычный программный комплекс при машинном переводе­­­­­­­­­­­­­­­ на основе моделирования грамматических категорий глаголов и морфологического анализа.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. V МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ ПО КОМПЬЮТЕРНОЙ ОБРАБОТКЕ ТЮРКСКИХ ЯЗЫКОВ «TURKLANG 2017» МОДЕЛЬ МОРФОЛОГИЧЕСКОГО АНАЛИЗА КЫРГЫЗСКОГО ЯЗЫКА. Т. Садыков, Б. Кочконбаева . стр. 135

Хотелось бы отметить, что по мнению Решетовой И.С. «помимо всего вышеуказанного необходимо учитывать и основные тенденции изменения коннотативного значения заимствований с национально-культурным компонентом в ходе их проникновения в переводимый язык с нейтрального на отрицательный, с отрицательного на нейтральный, с нейтрального на положительный. Анализируется семантическая деривация, которую претерпевает слово при изменении его оценочного компонента: расширение или сужение значения, метафорический или метонимический перенос, в процессе чего наблюдается динамика национально-культурного компонента.»2

**ТЕХНОЛОГИИ, МОДЕЛИ И СИСТЕМЫ**

1. Аннотирование узбекской грамматики

Грамматика состоит из двух частей: морфологии и синтаксиса.

**Части речи узбекского языка**

|  |  |  |
| --- | --- | --- |
| Самостоятельные части речи | Служебные части  речи | Отдельные группы слов |
| Существительное Наречие  Прилагательное  Местоимение  Глагол  Числительное | Союз  Междометие  Вспомогательные слова | Служебные слова  Модальные слова  Подражательные слова |

Грамматические значения, деривации, словоизменительные правила и формальных модели в морфологии рассматриваются как лингвистические процессы. Формальные морфологические модели являются результатом использования словосочетаний и отношений между ними в тексте. Формальные модели всегда существуют в синтагме.

Синтагма – семантико-синтаксическая единица, которая выражает некоторые унифицированные слова как значимую часть предложения. Лингвистическая база данных включает грамматику и словарь.

Любой, синтаксический анализ, содержит три основные составляющие:

1) Часть речи;

2) Член предложения;

3) Типы предложений.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. [КУЛЬТУРНОЕ НАСЛЕДИЕ ДРЕВНИХ И НАЦИОНАЛЬНЫХ ЯЗЫКОВ В ПЕРИОД ГЛОБАЛИЗАЦИИ](https://elibrary.ru/item.asp?id=37415307&selid=42861485)  
   МАТЕРИАЛЫ МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ. отв. ред. Е. А. Соболева. 2018. СОЦИОКУЛЬТУРНЫЙ ХАРАКТЕР ИЗМЕНЕНИЯ ОЦЕНОЧНОГО КОМПОНЕНТА СЕМАНТИКИ АНГЛОЯЗЫЧНЫХ ЗАИМСТВОВАНИЙ В ПРОЦЕССЕ АССИМИЛЯЦИИ (197-203) Издательство: [Армавирский государственный педагогический университет](https://elibrary.ru/publisher_about.asp?pubsid=15355) (Армавир)

Узбекский язык – это морфологически богатый язык с существительными, прилагательными и глаголами, изменяемыми по падежам, числам и другим формам слов. Данное свойство требует добавления морфологической информации в системы машинного перевода для устранения недостатка множества флективных форм. Так же параллельно с ней альтернативную информацию из английского языка. Для машинного перевода важно создать формальную грамматику узбекского языка во всех её формах и отражениях. Узбекский язык имеет агглютинативную морфологию с продуктивными флективными и деривационными суффиксами. Суффиксы могут добавляться последовательно, и одно слово может содержать много параметров, такие как притяжательность, множественное / единственное число, падеж, модальность и т. д.

Изменение по падежам – распространенная лингвистическая категория, присутствующая во многих языках мира. В литературе, посвященной формальному синтаксису, есть два основных подхода к выделению падежей. Первый подход в основном связан с работой Ноама Хомского, который рассматривает падеж как синтаксическое явление, известное в NLP; второй подход, предложенный в работе Алека Маранца, рассматривает падеж как предсинтактический, чисто морфологический феномен. 4

По мнению Н.Абдурахмановой «Существуют следующие деривационные модели узбекского языка:

W+A=>nok+zor

A+W=>be+foyda

W+W=>tez+yurar

W-W=>ota-ona

W W=>sotib olmoq

W-u/yu W=>Erta-yu kech

Из-за отсутствия грамматической информации для обработки естественного языка, осуществляется описание языка для лингвистической базы данных. Моделирование грамматических категорий для машинного перевода в узбекском языке производится в сравнении с английским языком. Английский и узбекский языки принадлежат разным языковым группам. Поэтому выделение уникальных свойств и различий обоих языков считается важным для морфологического анализа.»4

3. «Синтакси́ческие структу́ры» — лингвистическая работа, выпущенная Н. Хомским в 1957 году

4. V Международная конференция по компьютерной обработке тюркских языков «TURKLANG 2017» MODELING GRAMMATICAL CATEGORIES OF VERB IN UZBEK AS STAGE OF MORPHOLOGICAL ANALYSIS IN MACHINE TRANSLATION N. Abdurakhmonova стр. 155

Если рассмотреть данный процесс на примере глаголов узбекского языка можно убедиться, что процесс перевода – трудная работа из-за ментальных и концептуальных различий, которые существуют в разных языковых семьях, обществе и культурах. На качество перевода влияют, как лингвистические (неоднозначность, синонимы, паронимы, омонимы), так и экстралингвистические (психологические) факторы, а также культура и менталитет. Даже человек-переводчик сталкивается с теми же проблемами в процессе перевода, с которыми сталкиваются системы машинного перевода.

Когда перевод производится между родственными языками, перевод делать легче, но, когда это контексты из неродственных языков это делает машинный перевод мало эффективным, так как корни слов, окончания, предлоги могут иметь разное значение. Осложняется работа переводчика ещё и порядком построения предложений, способами выражений. При сопоставлении двух языков можно заметить, что два языка имеют одни и те же глагольные формы: 1. Императив.

2. Глаголы совершенной формы используются для прошедшего времени в узбекском языке.

3. Глаголы несовершенной формы используются для будущего времени в английском языке, но используются для выражения различных времен в узбекском языке (прошлого, настоящего и будущего) в сочетании с различными наклонениями и частицами.

4. В узбекском языке активные и пассивные причастия используются в меньшей степени чем в английском языке.

Глаголы имеют следующие грамматические категории:

Если посмотреть на агглютинативные языки, такие как финский, можно обнаружить, что морфосинтактические признаки систематически кодируются отдельными морфемами, которые расположены в линейном порядке. [5, 63]

Особую сложность при переводе составляют словоизменительные и синтаксические аффиксы. Если в узбекском языке существует более 6000 словарных слов и более 206 типов аффиксов и их вариаций частей речи, 130 из которых являются глагольными необходимо провести огромную работу чтобы ввести в машинный перевод всех вариантов в английском языке. А если ещё учесть синонимичные ряды и одного и второго языков, то можно понять, что в этом направлении необходимо проводить комплексные работы и исследования. Этот факт стал основным толчком для автора статьи, владеющего английским и узбекским языками, основами информационно технических навыков, опытом работы переводчика для начала фундаментального труда и научно-практических исследований.

5. Andreea-Rosalia Olteanu. A holistic approach to phrasal verbs, Editura Sfântul Ierarh Nicolae 2012, P 16.

Для сравнения можно привести основу «uchmoq», мы могли видеть некоторые примеры различных моделей глагольных структур:

1) Простой глагол – uchmoq (fly)

2) Составной глагол – uchib ketmoq (fly away)

3) Коллокация – ракета uchirmoq (fly the …)

4) Сочетание с глаголом – varrak uchirib bermoq (fly the kite to

smb.)

5) Сочетание с модальным словом – uchirish kerak (must fly)

6) Идиома – kapalagim uchib ketdi (be afraid)

При вводе текста морфологический анализатор должен правильно анализировать каждый сегмент в тексте. В противном случае, при переводе единиц текста возникают проблемы омонимии. Например, комбинация слов qo‘yib berdi используется во многих функциях как контекстная омонимия, как в следующих примерах:

U hujjatni stolga qo‘yib berdi-> He gave document as putting on the table.

U bolani hovlida o‘ynab olishiga qo‘yib berdi-> He let the boy play in the yard.

Direktor ko‘rsatilgan hujjatlarga darhol imzo qo‘yib berdi-> The director signed abruptly brought documents.

U bolalar o‘ynab olsin deb, sho‘x ashula qo‘yib berdi-> He played music so that to dance the children.

Целью изучения на данном этапе исследования является создание базы данных фразовых глаголов как аналитических моделей в англо-узбекском переводе.

Для построения системы машинного перевода с английского языка на узбекский язык должен быть указан размер словаря, который был сохранен в базе данных. На английском и узбекском языках имеются очень большие базы данных, включающие все лингвистические уровни, и они очень разные. Глагольная категория на английском языке – фразовый глагол. Так или иначе, фразовые глаголы на английском языке, как глагольное словосочетание на узбекском языке имеют свои особенности. Это проблема для структурных компонентов предложения. Фразовые глаголы считаются очень важной и часто встречающейся особенностью английского языка. Во-первых, они настолько распространены в повседневном разговоре, и иностранцы, которые хотят казаться естественными, когда говорят на английском языке, должны изучать грамматику, чтобы знать, как правильно их произносить. Во-вторых, привычка изобретать фразовые глаголы была источником большого обогащения языка. С помощью фразовых глаголов описывается наибольшее разнообразие человеческих действий и отношений [6, p 16]. А это значит, что глагольные конструкции английского языка очень сложны для анализа и когерентного описания в синхронных терминах.

В заключение необходимо отметить, что:

- лингвистические модели и семантические отношения каждой языковой единицы играют важную роль при создании баз данных для систем машинного перевода;

-из-за процессов глобализации все меняется;

-нет препятствий для унификации культурных и социальных отношений между людьми.

Поэтому понимание иностранных языка очень важно, мы не можем это не учитывать при составлении абсолютно нового максимально включающего в себя все конструкции двух языков, и наиболее удобное программное обеспечение. Сегодня результат машинного перевода, который появился в последней половине 20-го века, играет огромную роль в развитии наук и взаимообмена информаций.

**ЛИТЕРАТУРА**

1. Лютикова Е. А. Формальное Моделирование падежного варьирования: параметрический подход // Компьютерная лингвистика и интеллектуальные технологии По материалам ежегодной международной конференции «Диалог» (2016) Выпуск 15, C. 461.

2. Материалы международной научно-практической конференции. отв. ред. Е. А. Соболева. 2018. «[Культурное наследие древних и национальных языков в период глобализации](https://elibrary.ru/item.asp?id=37415307&selid=42861485)» «Социокультурный характер изменения оценочного компонента семантики англоязычных заимствований в процессе ассимиляции» (197-203) Издательство: [Армавирский государственный педагогический университет](https://elibrary.ru/publisher_about.asp?pubsid=15355) (Армавир)

3. «Синтакси́ческие структу́ры» — лингвистическая работа. Н. Хомски 1957г.

4. Материалы «V международной конференции по компьютерной обработке тюркских языков «turklang 2017» modeling grammatical categories of verb in uzbek as stage of morphological analysis in machine translation» N. Abdurakhmonova стр. 155

5. Brian Roark, Richard Sproat. Computational Approaches to Morphology and Syntax. Oxford University Press Inc., New York, 2007, P. 63.

6. Andreea-Rosalia Olteanu. A holistic approach to phrasal verbs, Editura Sfântul Ierarh Nicolae 2012, P 16.

**Интернет ресурсы.**

1.eLIBRARY ID: [42861485](https://elibrary.ru/item.asp?id=42861485)/ EDN: [AHDNHD](https://elibrary.ru/ahdnhd)

2.https://cyberleninka.ru/article/n/lingvofilosofiya-noama-homskogo-ot-kartezianskoy-traditsii-k-generativnoy-grammatike#

3.http://lectura.bibliotecadigitala.ro/Olteanu\_Adriana/A\_holistic\_approach\_to\_phrasal\_verbs.pdf