ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

МОДЕЛИРОВАНИЕ АНАЛИТИЧЕСКИХ ФОРМ ГЛАГОЛА В УЗБЕКСКОМ ЯЗЫКЕ КАК ЭТАП МОРФОЛОГИЧЕСКОГО АНАЛИЗА В МАШИННОМ ПЕРЕВОДЕ

Бобур Аллабердиев [1], **Шарипбай Алтынбек** [2], **Зафар Рахмонов** [1]
allaberdiyev_91@mail.ru, sharalt@mail.ru, zraxonov@inbox.ru
1. Национальный университет Узбекистана
2. Л.Н. Евразийский национальный университет имени Гумилева, Астана, Казахстан

АННОТАЦИЯ

В статье рассматривается значимая роль морфемного анализа для моделирования грамматических категорий частей речи на узбекском языке в машинном переводе. Выявление типов этапов морфологического анализа и общих парадигм, различий исходного и целевого языков. Глаголы в узбекском языке имеют свои особенности в отношении форм и аналитических характеристик. Существуют составные глаголы, словосочетания; глагольные словосочетания и их формы играют важную роль для вычислительной морфологии. В статье показано моделирование грамматических категорий, основанных на формах, ограничении синтаксических отношений и сочетаниях аффиксов в формах глаголов.

Ключевые слова: естественный язык, машинный перевод, аналитические формы, автоматическая морфология, обработка естественного языка

ABSTRACT

The article considers the significant role of morphemic analysis for modeling grammatical categories of parts of speech in the Uzbek language in machine translation. Identification of the types of stages of morphological analysis and common paradigms, differences between the source and target languages. Verbs in the Uzbek language have their own peculiarities in terms of forms and analytical characteristics. There are compound verbs, word combinations; verb phrases and their forms play an important role for computational morphology. The article shows the modeling of grammatical categories based on forms, the restriction of syntactic relations and combinations of affixes in the forms of verbs.

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

Keywords: natural language, machine translation, analytical forms, automatic morphology, natural language processing

ВВЕДЕНИЕ

Одна из очевидных заслуг информационных технологий заключается в том, что они заставляют развивать науку и положительно влиять на оптимизированную инфраструктуру отрасли сферы знаний, главным образом связываться с широкополосной связью по всему миру. Вообще говоря, системы каждой ветви общества реформировались после обретения независимости в нашей стране. Кроме того, став прогрессивными изменениями, образовательная система с тех лет прогрессировала одновременно. В частности, распространение в больших масштабах вычислительных технологий, а именно разоблачение возможность использования интернет-системы сделала открытой дверь мира для решения серьезных проблем в науке. Появились и продолжались большинство направлений наук, пересекающихся с компьютером.

В качестве яркого примера можно привести компьютерную лингвистику, появившуюся в 2000-е годы по инициативе профессора А.Пулатова, где она была создана в качестве первой лаборатории на факультете узбекской филологии Национального университета Узбекистана имени Мирзо Улугбека (НУУз). Основной целью было сформулировать идеологию концептуализации узбекской компьютерной лингвистики. В те годы некоторые темы исследовались студентами магистерского кружка по темам моделирования, машинного перевода, автоматической орфографии, создания информационного стиля узбекского языка. В направлении кандидат математических наук М.Хакимов (НУУз) применил множество исследований по математическому моделированию подходов к многоязычному машинному переводу. Количество прикладных работ, выполненных под его руководством. Были созданы некоторые справочники и учебные пособия, а именно “Компьютерная лингвистика” (А.Пулатов, 2011), “Основы компьютерной лингвистики” (А.Рахимов, 2011), “Лингвистические основы машинного перевода” (Н. Абдурахмонова, 2012) и др. Тем не менее, большинство реализованных работ были в теоретическом аспекте, поэтому никаких реальных программ, основанных на полной лингвистической базе данных, не были разработаны. В настоящее время компьютерная лингвистика как наука преподается в нескольких государственных университетах, а именно Андижанской, Наманганской, Ферганской, Хорезмской, Самаркандской, Бухарской областей

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

Узбекистана. Больше всего центром расследования для CL является Ташкентский государственный университет узбекского языка и Литературный университет имени Алишера Навои, который был создан 13 мая 2016 года указом первого президента И. А. Каримова. Несмотря на исследования флуктуаций, после декрета о CL можно считать, что он достиг пика внимания к проектированию и созданию программы курсов бакалавриата и магистратуры, несмотря на очень короткое время. Потому что в соответствии с этим указом были четко обозначены некоторые существенные вопросы: “...обеспечение надлежащего места нашего родного языка в мировой информационной системе Интернета, его вычислительного стиля, научно-методических пособий, связанных с автоматическим переводчиком и электронными словарями, подготовка прикладных рекомендаций и широкое внедрение результатов, успешно достигнутых на практике”. Мы должны охотно признать это уважительное отношение к узбекскому языку и потоку всякого рода исследований, которые оборачиваются только положительной стороной.

Узбекский язык – это язык великого Алишера Навои, который заложил богатую сокровищницу этого языка. Как известно, узбекский язык относится к тюркским языкам и, надо признать, имеет давнюю историю с изменчивыми позициями своей судьбы под воздействием различных факторов. Свои особенности узбекского языка среди других языков мы видим на каждом уровне языкознания. Например, как мы приводим пример, сохранилась гармония гласных в таких словах, как *üzüm, velâyet* в фонетике и сохранились национальные слова в турецком языке. Однако заимствований в узбекском языке больше, чем в турецком. Например, *management, budget, test* с английского, *стол, поезд, бухгалтер* с русского, *вазир, мактаб, маориф* с Арабский. С одной стороны, лексикология, рассматриваемая как очень динамичная система, имеет дело и с социальными и политическими ситуациями. То есть, если говорить об узбекском языке, то его графемная система также несколько раз менялась на протяжении веков. После установления независимости Узбекистана, а именно в 1989 году узбекский язык был принят в качестве государственного, постоянно проводились реформы по его совершенствованию. Одним из них являются орфографические правила, основанные латинским письмом, утвержденным в 1995 году. Одним из значимых вопросов компьютерной лингвистики в узбекском языке является создание вычислительного анализатора.

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

КРАТКИЙ ОБЗОР ГРАММАТИКИ УЗБЕКСКОГО ЯЗЫКА

Грамматика состоит из двух частей, а именно морфологии и синтаксиса.

Parts of speech of Uzbek:			
Dominant elements of a sentence		**Secondary elements of a sentence**	**Separated groups of the words**
Noun	Adverb	Conjunction	Interjections
Verb	Numeral	Auxiliary (Yuklama)	Imitative words
Adjective	Pronoun	Helping words (ko ʻmakchi)	Modal words

В качестве лингвистической процедуры рассматривается подход к репрезентации грамматических смыслов, деривации, правил словообразования и форм формальных моделей в морфологии. Морфологические формальные модели возникают при использовании словосочетаний и отношений друг к другу в тексте. Формальные модели всегда существуют в синтагме. Синтагма – это семантическая синтаксическая единица, выражающая некоторые единые слова как смысловую часть предложения. Лингвистическая база данных включает в себя грамматику и словарь. Как правило, синтаксический анализ реализуется по основным трем фазам в ходе автоматического процесса:

1) Части речи
2) Части предложения
3) Виды предложений

Узбекский язык-морфологически богатый язык с существительными, прилагательными и глаголами, изменяющимися по падежу, числу и формам слов. Это свойство требует введение морфологической информации внутри системы МТ, чтобы справиться с отсутствием многих флективных форм. Очень важно создать формальную грамматику узбекского языка для машинного перевода. Узбекский язык имеет агглютинативную морфологию с продуктивными флективными и деривационными суффиксами. Из-за того, что суффиксы могут добавляться последовательно, одно слово может передавать много информации, например притяжательную информацию, множественное/единственное число, падежная информация, настроение и т.д. Вариативность падежа-широко распространенное языковое явление.

В литературе, посвященной формальному синтаксису, можно найти два основных подхода к назначению падежа. Первый подход, который в основном связан с Намом Хомскийс рассматривает падеж как синтаксический феномен, который лицензирует NPS; второй подход, выдвинутый в работе Алека

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

Маранца, рассматривает падеж как пост синтаксический, чисто морфологический феномен. [1, 461]

W+A=>olma+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, он стремится ввести описательный язык для лингвистической базы данных.

Моделирование грамматических категорий в узбекском языке осуществляется в рамках английского языка для машинного перевода. Английский и узбекский языки принадлежат к разным языковым семьям. Поэтому нахождение уникальных особенностей и различий обоих языков считалось существенным делом для морфологического анализатора.

ФОРМАЛЬНАЯ МОДЕЛЬ ТРАДИЦИОННОЙ МОРФОЛОГИИ В МАШИННОМ ПЕРЕВОДЕ

Само собой разумеется, что процесс перевода является сложной работой из-за ментальных и концептуальных вопросов, которые существуют в разных языковых семьях, обществах и культурах человечества. Лингвистические (двусмысленность, синонимия) и экстралингвистические (психологические) факторы влияют на качество продукта перевода. Даже человек-переводчик сталкивается с такой помехой в процессе перевода, машинный перевод также сталкивается с этими проблемами. Как указано выше, независимо от того, связаны ли они или несвязанный язык берется за машинный перевод, между языками существует некая концептуальная идеология. Кемаль Алтынташ сравнивая крымско – татарского и Турецкие языки определил, что “порядок слов и обязанности слов в предложении в большинстве случаев сходны. Корни обычно похожи, но иногда они могут иметь разные значения в двух языках” [2, 30]. Машинный перевод между тюркскими языками проще, чем неродственные языки. Глаголы флексируют для числа, рода, лица и времени, и эти два языка имеют сложную и сходную структуру глаголов и система перегибов. Эти два языка имеют одни и те же глагольные формы:

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

1. Совершенная форма используется для прошедшего времени в узбекском языке

2. Имперфект используется для будущего времени в английском языке , но используется для различных времен в узбекском языке (прошедшее, настоящее и будущее) в координации с различными наклонениями и частицами.

3. Императив

4. Активные и пассивные причастия используются для настоящего времени в английском языке и в меньшей степени как глагол в узбекском.

Глагол содержит эти грамматические категории:

Lexeme	->	*o‘qidim*
aspect	->	simple
person	->	first
number	->	singular
voice	->	active
mood	->	indicative

Если посмотреть на агглютинативные языки, такие как финский, можно обнаружить, что морфосинтаксические признаки систематически кодируются отдельными морфемами, расположенными в определенных линейных порядках. [3, 63]

Существует более 50 аффиксов, придающих новый смысл флективным глаголам, чуть менее 30 словообразовательных аффиксов; синтаксические аффиксы охватывали более 30 форм. В целом глаголы качества в узбекском языке состоят из более чем 6000 слов в лексиконе. В узбекских языках насчитывается около 207 типов суффиксов (включая вариативность) частей речи, и 130 из них определяются как глаголы. Для того чтобы добавить окончания к основам каждого слова, необходимо разделить ту или иную часть речи на парадигмы [4, 12-17]

При морфологическом анализе в словаре приводятся основы слов с грамматической информацией и правилами. При сравнении ствола “uchmoq” мы могли бы увидеть несколько примеров различных моделей следующих глагольных структур:

1) Simple verb – ***uchmoq (fly)***

2) Compound verb – ***uchib ketmoq (fly away)***

3) Collocation – samalyot ***uchirmoq (fly the plane)***

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

4) Verbal word combination – varrak ***uchirib bermoq (fly the kite to smb.)***

5) Modal word combination – ***uchirish kerak (must fly)***

6) Idioms – ***kapalagi uchib ketmoq (be afraid)***

По мнению Юлдашева А. А. [5], в тюркских языках существует два типа аналитических форм глаголов:

- ***analytical forms of conditions;***
- ***analytical forms of modality;***
- ***analytical forms of other mood.***

Кроме того, морфологический анализатор должен правильно анализировать каждый сегмент текста. В противном случае возникают некоторые омонимические проблемы при переводе единиц в тексте. Например глагольное словосочетание **қўйиб берди** используется во многих функциях как омоним в контексте в следующих примерах:

У хужжатни столга **қўйиб берди** -> He gave document as putting on the table.

У болани ховлида ўйнаб олишига **қўйиб берди** -> He **let** the boy play in the yard.

Директор кўрсатилган худжатларга дархол имзо **қўйиб берди**-> The director **signed** abruptly brought documents.

У болалар ўйнаб олсин деб, шўх ашула **қўйиб берди**-> He **played music** so that to dance the children.

База данных и семантика глагольных словосочетаний изучены очень мало даже в узбекском языке. Учитывая, что все они являются основными глаголами, нет никаких чистых глагольных вспомогательных глаголов, которые они использовали бы независимо. Это около 30 типов глаголов, но они дают разные значения национальным глаголам. Они являются: **ber (ver), bil, bit (bitir), bor, boshla, boq, bo‘l, et, yoz, yot, kel, ket, ko‘r, ol, sol, tashla, tur, tush, chiq, yubor, yur, o‘l, o‘t, o‘tir, qara, qol, qo‘y.**

Существует два типа моделей словесных сочетаний:

- (MV+b/ib) HV
- (MV+a/y) HV

25 помогающих глаголов в сочетании с основными глаголами с аффиксом **-(i)b**:

berdi	**bo‘ldi**	**ko‘rdi (yoqib)**	**tushdi**	**o‘tdi**
Bildi	**(anglab) etdi**	**oldi**	**chiqdi**	**o‘tirdi**

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

O'qib	**bitirdi**	**yotdi**	**(aytib) soldi**	**yubordi**	**qaradi**
	bordi	**keldi**	**tashladi**	**yurdi**	**qoldi**
	boqdi	**ketdi**	**turdi**	**o'ldi**	**qo'ydi**

11 помогающих глаголов, соединенных аффиксами-**a/y** :

1) ber/ver (yoza berdi)
2) bil (topa bildi)
3) bor (o'zgara bordi)
4) boshla (yoza boshladi)
5) yoz (yiqila yozdi)
6) ket (gapira ketdi)
7) ko'r (ayta ko'rma)
8) ol (unuta oldi)
9) sol (kela solib)
10) tur (yoza tur)
11) qol (ayta qol)

16 помогающие глаголы не могут соединяться с основными глаголами через аффикс-a/y, они связывают друг друга только с помощью аффиксов **b/ib**: **bit (bitir), boq, bo'l, et, yot, kel, tashla, tush, chiq, yubor, yur, o'l, o't, o'tir, qara, qo'y.**

9 аффиксы могут присоединяться к обеим формам аффиксов (**b/ib** и **a/y**): **ber** (aytib ber, ayta ber), **bil** (yozib biladi (in dialect), yoza biladi), **bor** (o'qib bordi, o'qiy bordi), **ket** (isib ketdi, gapira ketdi), **ko'r** (aytib ko'r, ayta ko'rma), **ol** (yozib ol, yoza ol), **sol** (to'kib soldi, kela solib), **tur** (o'qib tur, yoza tur), **qol** (tugab qoldi, ayta qol).

Однако эти глаголы могут присоединяться к обоим аффиксам, но только они имеют разные значения: **yozib oldi (take a note)– yoza oldi (could write), aytib ko'r (try to tell) – ayta ko'rma (don't tell any more).** Иногда они имеют одно и то же значение: **og'irlashib bordi – og'irlasha bordi**.

Значения глагольных словосочетаний:

form	example
-(i)b ber	qo'yib ber, bilib ber
-a/y ber	ko'chaver, o'ylayver
-a/y ol/bil	tuzata oldi, foydalana bil
-(i)b bit/bitir	yonib bitgan, ekib bitir
-i)b, -a/y bor/kel	unutib bordi, o'zgartira bor
-a/y boshla	yura boshla, o'qiy boshla
-(i)b boq/ko'r/qara	o'qib boq, o'ylab qara, yozib ko'r
-(i)b bo'l	yozib bo'l, yuvib bo'l, qazib bo'l
-(i)b et	anglab etdi, pishib etmagan
-a/y yoz	qula yozdi, yorila yozdi
-(i)b yot/tur/o'tir/yur	o'qib yot, aytib tur, yozib o'tir
-(i)b ket	tarqab ketdi, isib ketdi, g'ovlab ketdi, maqtab ketdi
-a/y ket	o'qiy ketdi, gapira ketdi
-(i)b sol	aytib sol, yuragini to'kib soldi
-a sol	ola solib ot, kela solib boshla, tura solib tashlan
-(i)b tashla/yubor	o'qib tashla, to'qib tashla, haydab yubor
-(i)b chiq	o'qib chiq, so'rashib chiq, aylanib chiq
-(i)b tush	ag'darilib tushdi, yiqilib tushdi, yoqib tushdi, yarashib tushdi

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

-(i)b o‘l	mudrab o‘lyapman, surishtirib o‘ldi, sog‘inib o‘ldi, chanqab o‘ldi
-(i)b o‘t	bo‘lib o‘tdi, gapirib o‘t
-(i)b qol	tikilib qol, serrayib qoldi; kelib qoldi, so‘rab qol; jonlanib qoldi
-a/y qol	bora qol, yura qol, keta qol
-(i)b qo‘y	suyab qo‘y, ilib qo‘y; o‘ylantirib qo‘y, shoshirib qo‘y; qarab qo‘y, yo‘talib qo‘y

Помогающие глаголы иногда пишутся как укороченные формы или составные глаголы:– *aytib yubor – aytvor, bora ber – boraver, unuta olmadi – unutolmadi.*

В составе такого рода глаголов есть несколько глаголов: айтиб бериб тура кул, айтиб бра олмай голди. Каждый раз первая часть глаголов считается главным глаголом: Aytib (главный глагол) ber (помогающий глагол), aytib (главный глагол) berib tur (помогающий глагол).

Глагольные словосочетания похожи на составные глаголы, но только первый компонент глагола дает основное значение, другие помогают этому основному глаголу, в составном глаголе сохраняют оба компонента, сохраняют независимые значения. Мы видим три одинаковые модели:

1. (MV+PP) HV=> oshib tushmoq (Compound verb)
2. (MV+PP) HV=> oshib ketmoq (Verbal word combination)
3. (MV+PP) HV=> to‘pni oshirib uzatmoq (Collocation)

1) если убрать помогающий глагол из условных глаголов, то такое сочетание может несколько изменить смысл: **yozib bordi (write on) – yozdi (write), isib ketdi (warm up) – isidi (get warm)**.

Тем не менее, снятый помогающий глагол из составного глагола будет полностью менять смысл слова, потому что помогающий глагол участвует в образовании нового слова: **sotib ol (buy) – sot (sell)**; **ishlab chiqardi (produce) – ishladi (work);**

2) в тексте имеется более двух независимых единиц словосочетания: **Quvonib** (modifier) **so‘zladi** (predicate) – *he talked joyfully*; в глагольном словосочетании есть только один сказуемый: **O‘ylab qoldi** (predicate).

Аффиксы голоса и отрицательной формы могут быть добавлены как части глагольного словосочетания: **to‘xtatib qo‘yishdi; aytib qo‘yma, aytmay tur, aytmay turma;** к глаголам помощи присоединяются аффиксы времени, настроения, лица. Помимо этих синтаксических форм существуют одновременно в обеих частях глагола: **tamomlashdi-qo‘yishdi.**

В дискуссиях о морфологии принято говорить о флективных вариантах деривационной морфологии в узбекском языке, с точки зрения типов

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

признаков, которые каждый из них кодирует. Важным вопросом является моделирование грамматических форм в машинном переводе:

1) Моделирование вербальных словосочетаний;

2) Нахождение адекватного смысла глагольных словосочетаний в английском языке.

Это общие модели вербальных словосочетаний:

- MV HV->o‘qib berdi
- MV+HV->berolmadi<=> bera olmadi
- [MV]-[HV]->yozdi-qo‘ydi

Вот некоторые особенности моделирования:

MV-main verb (MV – такие главные глаголы сохраняют собственные национальные значения)

HV1-helping verb (HV1 – такие глаголы добавляются после аффиксов-b/-ib)

HV2-helping verb (HV1 – такие глаголы добавляются после аффиксов-a/-y)

- MV+HV1=>aytib berdi
- MV+HV2=>so‘zlay oldi

Глагольные словосочетания похожи на фразовые глаголы в английском языке тем, что некоторые предлоги или наречия придают дополнительные значения условным глаголам:

o‘qib bo‘l-	Mushtariy kitobni o‘qib bo‘ldi.
o‘qib chiq-	Mushtariy kitobni o‘qib chiqdi.
o‘qib tur-	Mushtariy kitobni o‘qib turdi.
o‘qib yubor-	Mushtariy kitobni o‘qib yubordi.
o‘qib tashla-	Mushtariy kitobni o‘qib tashladi.
o‘qib ol-	Mushtariy kitobni (qayta) o‘qib oldi.
o‘qib ko‘r-	Mushtariy kitobni o‘qib ko‘rdi.
o‘qib qo‘y-	Mushtariy kitobni (o‘zi uchun) o‘qib qo‘ydi.
o‘qib ber-	Mushtariy kitobni (ukasiga) o‘qib berdi.
o‘qib ket-	Mushtariy kitobni (to‘xtamay) o‘qib ketdi.

Тем не менее, сходство в обоих языках, есть один существенный угол фразовых глаголов, изменивших определенно значения в отличие от глагольных фразовых глаголов в узбекском языке. Кроме того, он не может сравнивать две категории как морфологическую единицу в силу собственной специфики языков и моделей вербального словосочетания в зависимости от того, в каком и как положении находятся:

- MV+HV=> ko‘rib qoldi
- MV+HV(1) + HV(2)=>ko‘rsatib bera oldi

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

- MV+HV(1) + HV(2) +HV(3)=>berib qo‘ya qoldi
- MV+HV(1) + HV(2) +HV(3) +HV(4)=>aytib berib qo‘ya qoldi
- MV+HV(1) + HV(1)=>o‘qib **tura tur**
- MV(1)+HV(1)=>**tura tur**
- MV(1)+HV(1) +HV(2)=>**turib tura** qolgin

Как мы указывали выше, некоторые модели [глагол + глагол] включающие в себя одни и те же корни, могут встречаться несколько раз и давать отдельные значения в тексте [6, 55].

Более того, они также выглядят как составные глаголы с внешней формой в соответствии со структурой:

- Oshib tushmoq->Compound verb (climb over)
- Oshib ketmoq->Verbal word combinations (rise up)

С другой стороны, помогающие глаголы также являются компонентом сложных глаголов и идиом:

- Nonushta qilmoq (have a breakfast) =>nonushta qilib berdi(Compound verb);
- Mashq qilmoq (do exercise) =>mashq qilib turdi (Compound verb);
- Kapalagi uchib ketmoq (to be afraid) => kapalagini uchirib yubordi (Idiom);

Морфологический анализатор должен правильно идентифицировать каждую единицу в соответствии с ее контекстуальным значением. В узбекском языке глагол имеет следующие предикативные формы:

Negative form-N {-ma| -mas| -may}
Conditional form-CF {-sa}
Tense-T {-a|-y|-yap|-moqda|-yotir…}
Predicative form-PF {-man|-san|-dir…}
Voice-VF {-t|-tir|-giz, |-kiz…}
Non-finite form-NF {-gan|-kan|-qan|-b|-ib…}
Person-P {-im|-ing|-k|-ngiz|-lar| -man| -san|-k|-ngiz}
Purpose-PS {-moqchi}

ПРИМЕР КОМБИНАЦИИ

- MV+ N+PF=> O‘qimayman
- MV+ N+ T+PF=> O‘qimadim
- MV+ CF +PF=> O‘qisam
- MV+ N +CF +PF=>O‘qimasam

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

- MV+ T+PF=>O‘qiyapman
- MV+VF=>O‘qittir
- MV+NF=>O‘qigan
- MV+PS+PF=>o‘qimoqchiman

Аффиксы могут быть добавлены в обоих компонентах глагольного словосочетания:

- (MV+N+NF)+HV=> ko‘rmay qoldi
- (MV+ NF)+(HV+ N)=> ko‘rib qolmadi
- (MV+N+NF)+(HV+ N)=> ko‘rmay qolmadi
- (MV+V+NF)+HV =>ko‘rsattirib qo‘ydi
- (MV+NF)+(HV+ PF)=>ko‘rib borsam
- (MV+NF)+(HV+ N+PF)=>ko‘rib bormasam
- (MV+NF)+(HV+ VF+N+T)=>yeb ko‘rgizmadi

Помимо перечисленных выше всех форм глагольного словосочетания существуют некоторые структуры глаголов ***edi, ekan, emish*** и ***bo`lmoq, hisoblanmoq, sanalmoq, deyilmoq.***

Эти связывающие глаголы (эти глаголы создают предикативные формы) формулируют предикативные формы глаголов, которые они имеют в следующих моделях:

a) N|Adj.|Num.|Pron.|Adv.+bo`lmoq=> Agar soat o‘n bo‘lsa, sizga qo‘ng‘iroq qilaman (If it is 10 o’clock, I shall call you);

b) yo`q/bor/oz/ko`p/zarur/lozim/kerak+bo`lmoq=> O‘ylagan orzularim bir pasta yo‘q bo‘ldi (My thinking dreams disappeared abruptly);

d) Infinitive+kerak/lozim/shart/darkor: xona tozalanishi kerak=> Xona tozalanishi kerak (The room needs cleaning.

-The room needs to be cleaned).

ГЛАГОЛЫ В УЗБЕКСКОМ ЯЗЫКЕ ДЛЯ СИСТЕМЫ МОРФОЛОГИЧЕСКОГО АНАЛИЗА

Автоматический морфологический анализ восходит к самым ранним работам в компьютерной лингвистике по машинному переводу в 1950 - е годы (Andron, 1962;Woyna, 1962; Bernard-Georges et al., 1962; Boussard and Berthaud, 1965; Vauquois, 1965; Schveiger and Mathe, 1965; Matthews, 1966; Brand et al., 1969; Hutchins, 2001). На протяжении многих лет было много приложений, включая стеммер Портера (Porter, 1980), широко используемый в приложениях поиска информации (Dolby et al., 1965; Attar et al., 1978; Choueka, 1983; Büttel et

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

al., 1986; Meya-Lloport, 1987; Choueka, 1990; Коскенниеми, 1984), проверка орфографии (Макилрой, 1982; Хэнкамер, 1986), ввод текста систем (Беккер, 1984; Абэ и соавт., 1986) и морфологического анализа текста-к - синтеза речи (Аллен и др., 1987; Церковь, 1986; Кокер соавт., 1990). Многие из этих более ранних приложений использовали довольно специальные подходы, включая жесткое кодирование большей части лингвистической информации в систему. Например, в системе, описанной в Coker et al. (1990), большая часть морфологического анализа опосредуется таблицами, закодированными в виде заголовочных файлов C, и правилами изменения орфографии, записанными в виде функций C [9, 100].

Нужен Менеджер слов в машинном переводе. Word Manager (WM) - это система для морфологических словарей. Морфологический словарь-это база данных, в которой лексические знания организованы в соответствии с морфологическими правилами. [10, 88]

С нашей точки зрения, морфологический анализ не должен ограничиваться не только разбором категорий, главным образом в машинном переводе. Что касается Родольфо Дельмонте, то в итальянском языке лингвистические категории представлены следующими типами [11, 4-5]:

- грамматические категории - производные от категоризации реальности на сущности - существительные -, события - глаголы и номиналы -, атрибуты - прилагательные, наречия и существительные;
- семантические категории, такие как отрицание, кванторы;
- категории дискурсивного уровня, такие как дейктика, определенность, союзы для координации и подчинения на пропозициональном уровне;
- синтаксические категории - кодирование арности предикатно - аргументных структур в том виде, в каком они интерпретируются в ситуациях;
- аспектуальные категории - кодирование внутренней временной структуры событий (как выражается как вербальными, так и девербальными номиналами);
- семантические концептуальные категории - классификация типов событий по отношению к (не)реальности, которую они кодируют.;
- селекционные ограничения - кодирование типичности участников событий в присущих им семантических признаках, представленных в онтологии или связанной энциклопедической базе данных сущностей и их семантических взаимосвязей.;
- грамматические ограничения - кодирование так называемых

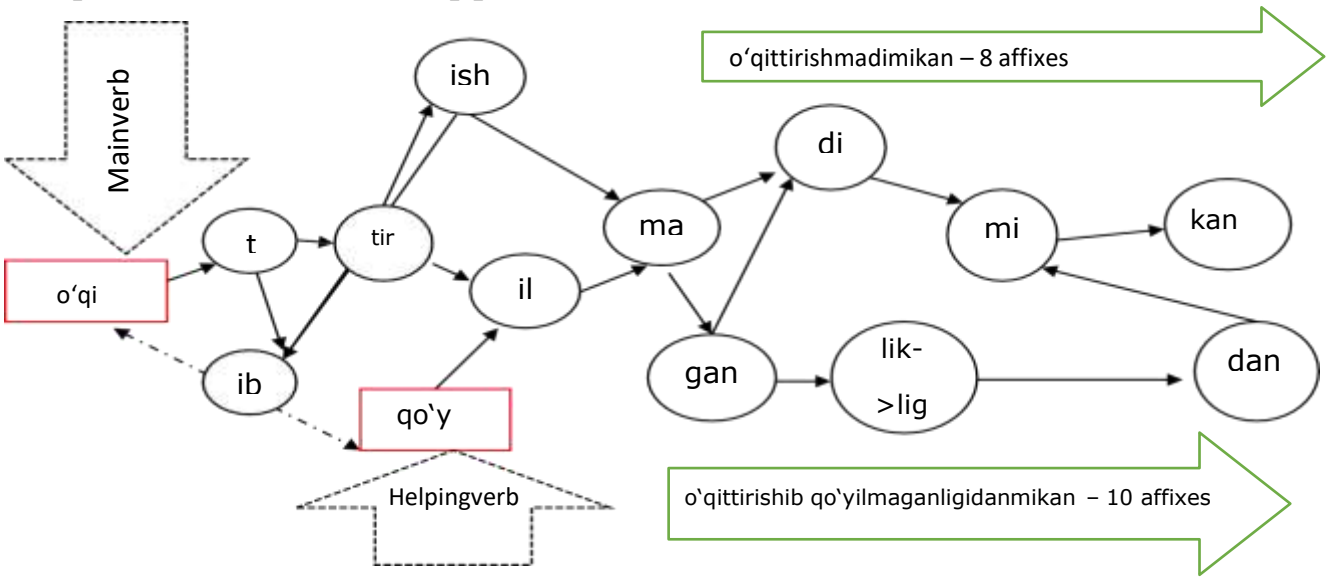
ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

синтаксических и анафорических ограничений связывания аргументов предикатов и зависимых предикатов только для пропозициональных аргументов.

Как мы видим, грамматические категории составляют основную часть морфологического анализа. Согласно некоторым работам, автоматический анализ имеет следующие виды [12, 65]:

- Анализ стебля в лексиконе
- Анализ словоформ в лексиконе
- Анализ с помощью логического подхода
- Анализ с помощью таблиц без лексики

Какой из типов подходит для узбекского языка, перейдем к следующей алгоритмической системе аффиксов глагола:



Здесь некоторые глаголы любят корень *<ko'n>*, но дают частично определения*: ko'nmoq<=>ko'nikmoq*

1. V->ko‘n (agree) -> U mening shartimga ko‘ndi (He agreed to my condition)
2. stem +ik=Verb ->ko‘nikmoq (used to)– Men shahar hayotiga ko‘nikdim (I am used to living in urban lifestyle)

Больше всего это должно быть направлено на глаголы в морфологическом анализе морфем, которые в том числе и в лемме. Записи лексикона маркируются с помощью простого алгоритма токенизатора длинного совпадения слева направо. Запись маркируется путем прохождения строки ввода, позиция за позицией, и поиска самых длинных символов, доступных с помощью очень простого жадного токенизатора. Если токенизатор инкрементный, он запоминает новые токены при анализе входных данных,

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

предполагая, что многосимвольные токены были объявлены заранее. Альтернативная, но менее эффективная стратегия состоит в том, чтобы определить все токены в отдельном проходе, чтобы составить входную строку с помощью токенизатора-преобразователя, реализующего жадное сопоставление слева направо или какую-либо другую стратегию для достижения желаемых разбиений [13]. Приведем еще один пример для анализа глагола в тексте: *Men hali ham* ***tuzalganim yo‘q****.–If automaton announce underlined word as predicate, the core problem is how it will be take apart morphemes correctly so that given proper meaning into another target language.*

Word manager ищет список аффиксов в базе данных. Отсюда будет очевидно, какие аффиксы являются правильной формой, к которой может быть добавлен глагол. Он имеет как список баз данных:

ID	Affixes	Function	Abbreviation
1.	ga	Case	C
2.	gan	Tense (Past)	T_{past}
3.	gan	Participle (Past)	PP
4.	l	Voice (Active)	$V_{act.}$
5.	l	Voice (Passive)	$V_{pass.}$
6.	a	Tense (Present)	$T_{pres.}$
7.	m	Person (first)	P_1
8.	i	Possessive pronoun (3-person)	$PossP_3$
9.	im	Possessive pronoun (1-person)	$PossP_1$

После проверки аффиксов проверяется модуль словообразования и сочетания аффиксов.

- Wstem+V+PP+PossP=>tuza+l+gan+im=> Passive voice
- Wstem +PP+PossP=>tuzal+gan+im=> Active voice

Здесь может возникнуть проблема, какая из них истина. Кроме того, принимая во внимание анализ этих данных, после этого процесса требуется семантический анализ. Предположим ради аргументации, что аффиксы дают различный смысл с различными функциональными позициями. Отрицательные формы глагола также считаются одной из важных парадигм в узбекском языке. Потому что вариативные формы также вызывают изменение смысла. Следующие модели предназначены для обозначения отрицательных форм в различных словосочетаниях:

I.MV+ma=>o‘qimadi (He did not read)

II. MV+may) HV=> O‘qimay qo‘ydi (He is not used to read)

III.MV (HV+ma)=>O‘qib qo‘ymadi (He did not read anymore)

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

iv.(MV+may) (HV+ma)=>O‘qimay qo‘ymadi – **positive meaning meaning**(Surely he read (in the past))

v. (MV+PP)+emas=>O‘qigan emas (He has never read)

vi.(MV+PP+Poss) yo‘q=>O‘qigani yo‘q (He did not read)

vii. Na MV{CV, MV, VC}=>na habar oldi | na o‘qidi | na berib ketdi (Neither he read)

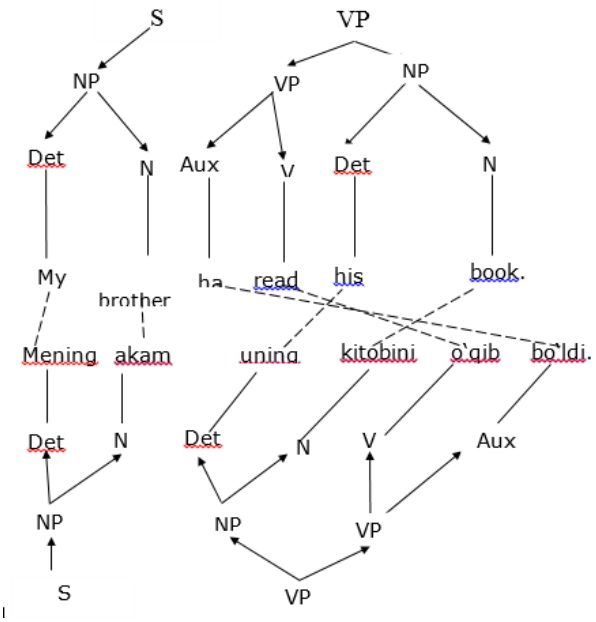
viii. (MV+mas) edi=>O‘qimas edi (He did not use to read)

ix. (MV+ma+gan)+ekan+P=>O‘qimagan ekanman (I did not read)

x. Na (MV+ma)=>Na o‘qimadi (He did not read)

Композиция регулярных отношений-это единственная наиболее общая вычислительная операция, которая может обрабатывать формальные устройства, найденные в морфологии естественного языка.

Оценено одно, что семантический полюс лексики и контекстуальное значение словоформ в системе машинного перевода очень необходимы для лингвистической базы данных. Из-за отсутствия слов и аффиксов сочетаний и не дано уместность, грамматические категории в исходном и целевом языках вызывают только электронные словари, но не машинный перевод. В тексте перевод на узбекский должен был звучать так: Мен барча ишларимни килиб боулдим. Давайте проанализируем подобный текст в контекстно свободной грамматике:



Как визуализируется модельное предложение, должно быть ясно, как части предложения связаны друг с другом в свою очередь на обоих языках. В частности, в каждом языке, показанном на диаграмме, нет некоторых

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

категорий. Соответствующие части речи и семантические полюсы слов должны учитываться и в машинном переводе.

ЗАКЛЮЧЕНИЕ

В целом, лингвистические модели и семантические отношения каждой единицы в машинном переводе играют важную роль для построения базы данных. Благодаря глобализационному процессу все начинает быстро меняться, кроме того, нет никаких препятствий для унификации культурных и социальных отношений между людьми. Таким образом, понимание другого иностранного языка стало очень важной вещью, которую мы не можем игнорировать. Сегодня результат машинного перевода, появившийся в середине прошлого века, повлиял на расширение исследований в этой сфере-это направление компьютерной лингвистики, а современные информационные технологии дают возможность использовать их в любых отраслях общественной жизни. Насколько нам известно, хороший машинный перевод является одним из инструментов торговли для процедуры перевода. Использование только грамматических моделей не может решить всех проблем в системе машинного перевода. Само собой разумеется, что подстановка записей в словари в соответствии с грамматической информацией помогает искать в лингвистической базе данных. Однако если он продолжает внимательно изучать не полную информацию о языке, искусственный интеллект по-прежнему остается единственным инструментарием переводчика в процессе перевода.

REFERENCES

1. Лютикова Е. А. Формальное Моделирование падежного варьирования: параметрический подход // Компьютерная лингвистика и интеллектуальные технологии По материалам ежегодной международной конференции «Диалог» (2016) Выпуск 15, С. 461.
2. Kemal Altıntaşh Turkish to Crimean Tatar machine translation system (a thesis submitted to the department of computer engineering and the institute of engineering and science of Bilkent university in partial fulfillment of the requirements for the degree of master of science) 2001, P.30
3. Matlatipov G., Vetulani Z. (2009) Representation of Uzbek Morphology in Prolog. In: Marciniak M., Mykowiecka A. (eds) Aspects of Natural Language Processing. Lecture Notes in Computer Science, vol 5070. Springer, Berlin, Heidelberg

ISSN: 2181-1385
Scientific Journal Impact Factor (SJIF) 2021: 5.723
DOI: 10.24411/2181-1385-2021-00716

4. Sanatbek Matlatipov, Ualsher Tukeyev, Mersaid Aripov Towards the Uzbek Language Endings as a Language Resource.
5. Brian Roark, Richard Sproat. Computational Approaches to Morphology and Syntax. Oxford University Press Inc., New York, 2007, P. 63.
6. N.Abdurakhmonova, The bases of automatic morphological analysis for machine translation, Известия Кыргызский государственный технический университет им. И.Раззокова теоритической и прикладной научно-технический журнал, 2016 № 2 (38)], С. 12-17.
7. Yuldashev A.A. Аналитические формы глагола в тюркских языках. Москва, – Наука, 1965.
8. N.Abdurakhmonova. O‘zakdosh fe’llarning ketma-ket qo‘llanilishiga doir – Tilshunoslikka ilk qadam (to‘plam III), T., 2007, B. 55-59.
9. Andreea-Rosalia Olteanu. A holistic approach to phrasal verbs, Editura Sfântul Ierarh Nicolae 2012, P 16.
10. Yorick Wilks. Machine translation. Its scope and limits. Spring science+Business Media LLC. 2009 UK, P 139.
11. Brian Roark and Richard Sproat. Computational Approaches to Morphology and Syntax, 2007, OXFORD, P.100.
12. State of the Art in Computational Morphology, Workshop on Systems and Frameworks for Computational Morphology, SFCM 2009 Zurich, Switzerland, September 4, 2009 Proceedings, P. 88.
13. Rodolfe D. Computational Linguistic Text Processing: Lexicon, Grammar, Parsing and Anaphora Resolution. Nova Science Publishers, Inc. New York, 2008, P. 4-5.
14. Марчук Ю. Компьютерная лингвистика. Москва, 2006 С. 65.
15. State of the Art in Computational Morphology Workshop on Systems and Frameworks for Computational Morphology, SFCM 2009 Zurich, Switzerland, September 4, 2009, Proceedings, P. 31.
16. Дрейзин Ф.А. Об алгоритмизации составление алгоритма анализа языка. (на примере морфологии агглютинативного узбекского языка)//Научные труди Ташкентского университета, 1961, вып. 189.Матем.Науки, кн.С.121.
17. Абдурахмонова Н.З. Машина таржимасининг лингвистик асослари. Тошкент, «Академнашр», 2012 Б.