

Особенности машинного морфологического анализа и синтеза глаголов таджикского языка

Н. Ш. Мадибрагимов

Аннотация — На сегодняшний день компьютерная лингвистика таджикского языка находится на истоке своего развития. С целью развить эту сферу, автор данного исследования разрабатывает проект формализации словоизменений таджикского языка для машинного морфологического анализа и синтеза, в рамках которого требуется классифицировать таджикские слова по типам формообразования. Уже классифицированы слова существительные, прилагательные и местоимения. В данной статье автор предлагает классификацию глаголов таджикского языка. В статье выделены 9 типов формообразования глаголов и 4 подтипа для 181 слова. Для представленных типов и подтипов показаны их отличительные особенности. Данная классификация является продолжением исследований автора, которые включают в себя классификацию существительных, прилагательных и местоимений таджикского языка. Автором разработана программа в виде Интернет-приложения для реализации генерации форм слов таджикского языка. Это приложение работает на основе лингвистической базы, заполненной в результате текущих исследований, приведенных в данной статье. Классификация других частей речи таджикского языка продолжается.

Ключевые слова—компьютерная лингвистика, автоматическая обработка текста, таджикский язык, морфология таджикского языка, модель формообразования, генерация и определение форм слов, Интернет-приложение.

I. ВВЕДЕНИЕ

Общей целью исследования является формализация образования форм слов таджикского языка на основе универсальной модели формообразования [3]. На данный момент уже классифицированы существительные, прилагательные и местоимения таджикского языка. Результаты этих классификаций были представлены в статьях [1] и [2]. В этой статье рассказывается о продолжении исследования таджикских слов на классификации по типам формообразования, об исследовании глаголов таджикского языка. Описываются методы и принципы формообразования глаголов таджикского языка, с точки зрения метода генерации и определения форм слов с помощью представления формообразования последовательностью преобразований [3].

Мадибрагимов Навруз Шавкатович, Рязанский государственный радиотехнический университет имени В.Ф. Уткина (РГРТУ), 390005, Российская Федерация, Рязань, Гагарина, 59/1. Аспирант; Рязанский государственный медицинский университет имени академика И. П. Павлова (РязГМУ), 390026, Российская Федерация, Рязань, ул. Высоковольная, 9. Ассистент кафедры математики, физики и медицинской информатики РязГМУ (e-mail: navruzmadibragimov@gmail.com)

II. ОБЩИЕ СВЕДЕНИЯ О ФОРМООБРАЗОВАНИИ ГЛАГОЛА В ТАДЖИКСКОМ ЯЗЫКЕ

Глагол в таджикском языке является сложной категорией слова, не только по значению, но и по особенностям формообразования, относительно от других частей речи [4]. Глаголы меняются, по лицу, числу (единственное или множественное) и по времени: *гуфтам* (я сказал) – I-лицо, *гуфтӣ* (ты сказал) – II-лицо, *гуфт* (он сказал) – III-лицо; *мехонам* (читаю) – настоящее-будущее время, *мехондам* (читал) – прошедшее простое время; *хонда истодаам* (читаю сейчас) – настоящее время, *хондаам* (читал) – прошедшее время; *ман мехонам* (я читаю) – единственное число, *мо мехонем* (мы читаем) – множественное число.

Морфемы, которые в предложении указывают на субъект, называются местоименными суффиксами (*бандакчонишинҳо*) или энклитикой [4][5]. Местоименные суффиксы показывают взаимосвязь между субъектом (подлежащим) и предикатом (сказуемым). Только субъекты, выраженные в глаголе с энклитикой, могут быть либо в единственном, либо во множественном числе.

Таблица 1. Местоименные суффиксы в глаголе

лицо	единственное число	множественное число
I	-ам (ям)	-ем
II	-ӣ (ед ¹)	-ед
III	-ад (яд)	-анд (янд)

В таджикском языке исходной формой глагола, которая приводится в словарях, является инфинитив или неопределенная форма глагола [7]. Инфинитив или по-таджикски **масдар**, не указывает ни времени, ни лица, ни числа: *омадан* (прийти), *хондан* (учиться), *гурифтан* (брать). Любой инфинитив глагола таджикского языка оканчивается на *-тан* или *-дан*.

Во всех трех временах непроизшествие, невыполнение или отрицание в глаголах, обозначается образованием формы с помощью частицы *на-*, которая добавляется вначале глагола: *нагуфтан* (не сказать).

III. ОСНОВЫ ГЛАГОЛА

Глагол таджикского языка имеет две основы, которые служат базой для образования всех глагольных форм:

¹ В таджикском языке форма II-лица множественного числа, может употребляться также в единственном числе, обозначая уважительную или почтительную форму. Здесь можно увидеть аналогию с русским языком: *шумо мехонед* – вы читаете.

- основа прошедшего времени (ОПВ);
- основа настоящего времени (ОНВ).

Основа прошедшего времени определяется путем отбрасывания суффикса *-ан* от инфинитива. ОПВ по звучанию совпадает с формой 3-лица единственного числа простого прошедшего времени (табл. 2).

Таблица 2. Примеры

инфинитив	Основа прошедшего времени	Форма 3-лица ед.ч., простое прош. вр.
<i>омадан</i> (прийти)	<i>омад</i>	<i>омад</i> (пришел)
<i>кардан</i> (делать)	<i>кард</i>	<i>кард</i> (сделал)
<i>рондан</i> (водить)	<i>ронд</i>	<i>ронд</i> (водил)

Глагольные основы настоящего времени приводятся в словарях при инфинитиве.

По способу образования ОНВ глаголы делятся на правильные и неправильные. Это разделение условное, нужно только для нахождения ОНВ. У «правильных» ОНВ отыскивается по определенному правилу, а у «неправильных» ОНВ существуют свои. Правильные глаголы по способу нахождения ОНВ, делятся на следующие группы [6]:

1. Глаголы, инфинитив которых заканчивается на *-рдан* или *-ндан*, основы настоящего времени которых можно получить, если отбросить от инфинитива три последних буквы *-дан*:

Таблица 3. Примеры

Инфинитив	Основа настоящего времени
<i>хондан</i> (читать, учиться)	<i>хон</i>
<i>мондан</i> (оставлять)	<i>мон</i>
<i>хӯрдан</i> (кушать)	<i>хӯр</i>
<i>кандан</i> (рыть)	<i>кан</i>
<i>шумурдан</i> (считать)	<i>шумур</i>

2. Глаголы, инфинитив которых заканчивается на *-одан*, *-идан* или *-истан*, которых нужно отбросить для нахождения ОНВ:

Таблица 4. Примеры

Инфинитив	Основа настоящего времени
<i>истодан</i> (стоять)	<i>ист</i>
<i>молдан</i> (мазать)	<i>мол</i>
<i>коридан</i> (сеять)	<i>кор</i>
<i>шӯридан</i> (возмущаться)	<i>шӯр</i>
<i>донистан</i> (знат)	<i>дон</i>

Сюда входят глаголы, которые в инфинитиве оканчиваются на *-хтан*, *-удан*, *-штан*, *-фтан*. Для нахождения ОНВ, три последних буквы нужно отбрасывать, а четвертая заменяется следующим образом:

х на *з*,
ш на *р*,
у на *о*,
ф на *б/в*,

с на н/нд

Таблица 5. Примеры

Инфинитив	Основа настоящего времени
<i>омӯхтан</i> (изучать)	<i>омӯз</i> (учи)
<i>доштан</i> (иметь)	<i>дор</i> (держи)
<i>намудан</i> (казаться)	<i>намо</i> (показывай)
<i>гузоштан</i> (класть)	<i>гузор</i> (клади)
<i>рӯфтан</i> (подметать)	<i>рӯб</i> (подметай)
<i>фармудан</i> (приказывать)	<i>фармо</i> (прикажи)
<i>шикастан</i> (ломать)	<i>шикан</i> (ломай)
<i>кӯфтан</i> (молотить)	<i>кӯб</i> (молоти)
<i>бастан</i> (связывать)	<i>банд</i> (свяжи)

Неправильных глаголов в таджикском языке немного – около 30 штук.

В нашем исследовании в базу слов были добавлены инфинитивы глаголов, также их ОПВ и ОНВ: как для неправильных, так и для правильных.

Таблица 6. Примеры некоторых неправильных глаголов:

Инфинитив	Основа настоящего времени
<i>гаштан</i> (ходить, возвращаться)	<i>гард</i>
<i>кардан</i> (делать)	<i>кун</i>
<i>пухтан</i> (варить)	<i>паз</i>
<i>шудан</i> (становиться)	<i>становиться</i>
<i>омадан</i> (приходить)	<i>ой, о</i>

IV. АОРИСТ

В таджикской грамматике спрягаемая форма глагола, состоящая из основы настоящего времени и местоименных суффиксов называется *аористом*, которые совпадают с окончаниями настоящего-будущего времени [6].

Таблица 7. Спряжение аориста на примере глагола *нави-стан* (писать)

лицо	единственное число	множественное число
I	<i>нависам</i>	<i>нависем</i>
II	<i>нависӣ</i>	<i>нависед</i>
III	<i>нависад</i>	<i>нависанд</i>

Аорист совпадает с формами условно-сослагательного и принудительного наклонения. Одно отличие, для получения 2-лица единственного числа принудительного наклонения постфикс не требуется, когда у аориста постфикс *-ӣ*: *хон* (читай) – *хонӣ* (читаешь).

V. ФОРМЫ ГЛАГОЛЫ

Глаголы в таджикском языке делятся на простые, сложные и составные. Простые глаголы имеют один корень: *Мо китоб мехонем* (Мы читаем книгу). Сложные глаголы образуются с помощью префиксов *бар* и *дар*. Например, *бардоштан* (поднимать), *даргурифтан* (вспыхивать, загораться) [5]. Спряжение сложных глаголов отличается от простых тем, что в настоящем-будущем времени префикс *ме-* прибавляется не к пре-

фиксу сложного глагола, а к самому глагола: *дармеги-ронам* (включу) [6].

Составные глаголы имеют больше одного корня, т.е. формируются из нескольких корней, большинство которых образуются от существительных и глагола: *Онҳо саёҳат мекарданд ва дам мегирифтанд* (Они путешествовали и отдыхали). Еще несколько примеров составных глаголов:

пир шудан – стареть,
фарбех шудан – толстеть,
занг задан – звонить,
бу гирифтан – унюхать,
пай бурдан – догадаться,
нафас кашидан – дышать,
фаромуш кардан – забывать,
орзу намудан – пожелать и т.д.

Это все похоже на аналог словосочетаний или глаголов английского языка:

to get tired – монда шудан (уставать),
to get old – пир шудан (стареть),
to take care – бохабар шудан (заботиться),
to get hungry – гурусна шудан (проголодаться).

При формообразовании составных глаголов, видоизменяется только глагольная часть, а вспомогательная часть остается неизменной. Поэтому в нашей работе составные глаголы не рассматриваются. Глагольная часть составных глаголов или вспомогательные глаголы, исследованы как простые глаголы.

VI. НАКЛОНЕНИЕ

В глаголе воздействие и действие излагаются, либо по повествовательному способу, либо со значениями условия, повеления и предположения. По этому способу тоже, глаголы могут принимать различные формы. Для обозначения таких формообразований, в глаголе используется понятие наклонения (*сига*). В таджикском языке существует следующие наклонения [8]:

- 1) изъявительное (*сигаи хабарӣ*)
- 2) повелительное (*сигаи амрӣ*)
- 3) условное-сослагательное (*сигаи шартӣ-хоҳишмандӣ*)
- 4) предположительное (*сигаи эҳтимолӣ*)

Изъявительное наклонение обозначает реальное действие в одном из трех времен: настоящем, прошедшем или будущем. Например, *хондам* – читал, *мехонам* – читаю, *хоҳам хонд* – буду читать.

Повелительное наклонение обозначает какое-нибудь приказание, желание или требование (табл. 8).

Таблица 8. Примеры

лицо	единственное число	множественное число
I	<i>хонам</i>	<i>хонем</i>
II	<i>хон</i>	<i>хонед</i>
III	<i>хонад</i>	<i>хонанд</i>

Условное-сослагательное наклонение обозначает действие, которое должно выполняться с условием выполнения какого-нибудь другого действия: *Агар хонӣ, мефаҳмӣ* (Узнаешь, если прочтешь).

Предположительное наклонение выражает предположе-

ние о совершении какого-либо действия и могут выражаться во всех трех временах: *Муаллим омадагист* (Наверное, учитель пришел). *Ту ҳаҷм медонистагистӣ* (Наверное, ты тоже знаешь).

VII. ВСПОМОГАТЕЛЬНЫЕ ГЛАГОЛЫ

Глагол *ҳаст* (быть) является недостаточным, существительным глаголом, соответствует английскому глаголу to be. Он спрягается в формы всех лиц обоих чисел. Отрицательная формы – *нест*: *Ман муаллим ҳастам* (Я являюсь учителем); *Ман муаллим нестам* (Я не являюсь учителем).

Также существует глагол-связка *будан* (также в значении «быть»), но в отличие от *ҳаст*, имеет прошедшее время (табл. 9).

Таблица 9. Примеры

В настоящем времени	В прошедшем времени
Ман муаллим мебошам	Ман муаллим будам

Отрицательная форма образуется добавлением префикса *на-*: *намебошам*, *набудам*.

VIII. ВРЕМЕНА

В таджикском языке по инфинитиву нельзя определить совершенный и несовершенный вид глагола, как например, в русском языке. Но видовые значения глагола передаются с помощью весьма развитой системой времён глагола [6].

Таблица 10. Времена глаголов в таджикском языке.

название	пример
замони ҳозира-оянда (настояще-будущее)	<i>ман менависам</i> – я пишу; я напишу
замони ҳозираи муайян (настояще-определённое)	<i>ман навишта истодаам</i> – я пишу сейчас
замони гузаштаи оддӣ (прошедшее простое)	<i>ман навиштам</i> – я написал (обычно однократно)
замони гузаштаи наклӣ (основной перфект)	<i>ман навиштаам</i> – я написал (повествовательная форма; прошло, результат сохранился)
замони гузаштаи ҳикоягӣ (прошедшее длительное)	<i>ман менавиштам</i> – я писал
замони гузаштаи ҳикоягии наклӣ (перфект длительный)	<i>ман менавиштаам</i> – наверное, я писал
замони гузаштаи дур (преждепрошедшее)	<i>ман навишта будам</i> – я написал (давно)
замони гузаштаи дури наклӣ (преждепрошедший перфект)	<i>ман навишта будаам</i> – оказывается, я написал
замони гузаштаи давомдор (прошедшее определённое)	<i>ман навишта истода будам</i> – я писал, когда ...
замони гузаштаи муайяни наклӣ (перфект определённый)	<i>ман навишта истода будаам</i> – оказывается, я писал, когда ...
замони ояндаи таркибӣ (будущее составное)	<i>ман хоҳам навишт</i> – я буду писать

В данном исследовании изучены времена настоящего-будущее, прошедшее простое, основной перфект, прошедшее длительное, перфект длительный. Остальные времена глаголов образуются с помощью другой части речи – деепричастия, которая будет рассмотрена в следующих исследованиях.

IX. ЗАЛОГИ В ГЛАГОЛЕ

В таджикском глаголе есть действительный, страдательный, также понудительный залог. Действительный залог объясняет действие субъекта, распространяющее на объект, а страдательный залог, наоборот, когда субъект обозначает объект, на которого переходит действие другого предмета [6]. Рассмотренные выше методы формообразования глаголов, относятся к действительному залогу. Форма страдательного залога по виду является составным глаголом, и образуется с помощью сочетания деепричастия прошедшего времени и спряжением вспомогательного глагола *шудан*, иногда *гардидан* (делаться, становиться) [6]: *сохтан* (строить), *сохта шудан* (строиться). Так как в нашей работе рассматриваются спрягаемая часть составных глаголов, страдательный залог не представляется.

Понудительный залог или каузатив обозначает действие, совершённое лицом, указанным в глаголе, по воле другого лица [4]. Инфинитив каузатива образуется с помощью добавления к основе настоящего времени действительного залога суффикса *-ондан* (*-ёндан*) или *-онидан* (*-ёнидан*): *шиштан* (сидеть) – *шинондан* (посадить), *рӯидан* (расти) – *рӯёндан* (выращивать), *харондан* – *принудить купить*. Эти формы можно считать понудительным залогом 1-степени.

Понудительный залог также имеет форму 2-степени, которая образуется путем добавления суффикса *-ондан*, *-ёндан*, к основе настоящего времени понудительного залога. Можно отметить, что возможно получение 2-степени понудительного залога, добавлением суффикса *-онондан* (*-ононидан*), *-ёнондан* (*-ёнонидан*), к основе настоящего времени действительного залога.

Таблица 11. Пример понудительного залога

действительный залог	1-степень	2-степень
<i>фаҳмидан</i> – понять	<i>фаҳмондан</i> – объяснять	<i>фаҳмонондан</i> – объяснять (через кого-то)

X. НЕДОСТАТОЧНЫЕ ИЛИ БЕЗЛИЧНЫЕ ГЛАГОЛЫ

Глагол *боистан* (долженствовать) имеет только форму 3-лица единственного числа аориста *бояд* (должен) и прошедшего времени *мебоист* (должен был) [6]. У глагола *шоистан* (приличествовать) есть форма только 3-лица единственного числа аориста *шояд* (авось, наверное). Это слово почти утратило силу как глагол.

XI. КЛАССИФИКАЦИЯ СЛОВ ТАДЖИКСКОГО ЯЗЫКА ПО ТИПАМ ФОРМООБРАЗОВАНИЯ

Также как и для предыдущих исследований [1], [2], классификация глаголов таджикского языка основывается на результатах исследований Г.М. Довудова [9]. Г.М. Довудов в своей работе, исследуя префиксы и постфиксы частей речи таджикского языка, для глагола выявил 1046 аффиксов и разделил их на *словоизменительный*, *словообразовательный*, *словосочетательный* типы. Так как мы рассматриваем словоизменительные и словосочетательные аффиксы, из общего количество мы выделили 56 аффиксов для глагола.

Для анализа и классификации, были рассмотрены 180

глаголов неопределённой формы таджикского языка, из словаря учебника [8], а также еще несколько инфинитивов от автора статьи для полноты исследования. В словарь были добавлены слова, относящиеся к приставочным и сложным глаголам, которые отсутствуют в [8], для достоверности классификации таких глаголов.

Классификация глаголов по типам формообразования, была осуществлена похожим образом, как и проведенная классификация существительных, прилагательных и местоимений [1][2]. Из словаря [8] в алфавитном порядке рассматривались глагол, его ОНВ и ОПВ, и описывались все цепочки преобразований согласно модели формообразования [3]. Исследуемое слово сопоставлялось с выделенными типами формообразования. Если цепочки преобразований слова и типа совпадали, то слово относилось к этому типу. Иначе создавался новый тип формообразования и исследуемое слово относилось к нему.

XII. КЛАССИФИКАЦИЯ ГЛАГОЛОВ ТАДЖИКСКОГО ЯЗЫКА ПО ТИПАМ ФОРМООБРАЗОВАНИЯ

Как было отмечено, морфологические преобразования глагола таджикского языка базируются на двух основах глагола: ОНВ и ОПВ. Таким образом, при преобразовании инфинитива и для получения аориста, к ОНВ добавляются личные окончания аориста или местоименные суффиксы. Для временных преобразований постфиксы времен добавляются к ОПВ.

В таджикском языке личные окончания для аориста бывают двух видов, которые также совпадают с окончаниями настояще-будущего времени [6]:

1-й вид – постфиксы *+ам*, *+й*, *+ад*, *+ем*, *+ед*, *+анд*. Эти окончания чаще всего применяются к глаголам с окончанием ОНВ на согласную букву;

2-й вид – постфиксы *+ям*, *+й*, *+яд*, *+ем*, *+ед*, *+янд*, которые используются для глаголов с окончанием ОНВ на гласную букву.

После классификации формообразования глаголов были выделены следующие типы и подтипы формообразования глаголов таджикского языка:

1. F 1. Основы настоящего времени, заканчивающиеся на согласные буквы *б*, *в*, *г*, *д*, *ж*, *з*, *к*, *қ*, *л*, *м*, *н*, *п*, *р*, *с*, *т*, *ф*, *х*, *ҷ*, *ч*, *ш*. Этот тип характеризуется тем, что при морфологическом преобразовании и для получения аориста, к ОНВ добавляются личные окончания аориста из 1-го вида постфиксов (*+ам*, *+й*, *+ад*, *+ем*, *+ед*, *+анд*).

Примеры: *хондан* (читать), ОНВ – *хон*, аорист – *хонам*, *мехонам* (читаю), *хонй*, *мехонй* (читаешь); *донистан* (знать), ОНВ – *дон*, аорист – *донам* (знаю), *донй* (знаешь), *донад* (знает) и т.д.

Для создания каузативов двух степеней используются конструкции ОНВ+*он(и)дан*, ОНВ+*онон(и)дан* соответственно: *хонондан* (принудить читать), *хононондан* (принудить читать через кого-то). Спряжение по временам, где не используется деепричастие, а также образование формы предположительного наклонения, используют единые правила для всех типов.

2. F 2. ОНВ, заканчивающиеся на гласные буквы *ё*, *о*, *у*, *ӯ*. Аорист получается путем добавления местоимен-

- ных суффиксов из 2-го вида (+ям, +й, +яд, +ем, +ед, +янд).
- Примеры: *партофтан* (бросить), ОНВ – *парто*, аорист – *партоям* (брошу), *партой* (бросишь), *партояд* (бросит); *шустан* (мыть) – ОНВ *шӯ*, аорист – *шӯям* (мою), *шӯй* (моешь), *шӯяд* (моет) и т.д.
- Для создания каузативов двух степеней используются конструкции ОНВ+ён(и)дан, ОНВ+ёнон(и)дан соответственно: *партоёндан* (заставить бросить), *шӯёндан* (заставить мыть), *партоёнондан* (заставить бросить через кого-то), *шӯёнондан* (заставить мыть через кого-то).
- Некоторые ОНВ, имеют 2 вида, второй вариант можно отнести к типу F 3, например *чустан* (искать), ОНВ *чӯ* и *чӯй*, аорист – *чӯям* (ищу), *чӯй* (ищешь) и т.д. В нашей базе приведен только 1 вариант и слово отнесено к типу F 2.
3. F 3. ОНВ, заканчивающиеся на гласные букву «й». При создании аориста, буква «й» в конце удаляется. Личные окончания применяются по принципу типа F 2, т.е. местоименные суффиксы из 2-го вида (+ям, +й, +яд, +ем, +ед, +янд).
- Примеры: *поидан* (стеречь), ОНВ – *пой*, аорист – *поям* (стерегу), *пой* (стережешь), *пояд* (стережет) и т.д.; *зоидан* (родить) – ОНВ *зой*, аорист – *зоям* (рожу), *зой* (родишь), *зояд* (родит) и т.д.
- Для создания каузативов двух степеней используются конструкции ОНВ+ён(и)дан, ОНВ+ёнон(и)дан соответственно: *поёндан* (заставить стеречь), *поёнондан* (заставить стеречь через кого-то).
4. F 4. Сложные глаголы, образованные с помощью префиксов, такие как *бозгаиштан* (возвращаться), *бархӯрдан* (биться, врезаться), *даргирифтан* (зажигать) и другие. При спряжении сложных глаголов в настояще-будущем времени, префиксы прибавляются не к префиксу, входящему в состав производного глагола, а собственно, к глаголу [8]. Например, *дармегиронем* (зажигаем), *бознамегардад* (не вернется). При создании аориста, личные окончания, применяются по принципу типа F 1.
- Примеры: *бархӯрдан*, ОНВ – *бархӯр*, аорист – *бархӯрам*, *бармехӯрам* (врежусь), *барнамехӯрад* (не врежется); *бозгаиштан*, ОНВ – *бозгард*, аорист – *бозгардам*, *бозмегардам*.
- Для создания каузативов двух степеней используются конструкции ОНВ+он(и)дан, ОНВ+онон(и)дан соответственно: *даргиринондам* (принудил зажигать). Исключением являются сложные глаголы, после префиксов которых идет гласная: *баромадан* (выйти), *даромадан* (войти), *фуромадан* (спуститься) и другие. Они спрягаются как простые глаголы.
5. F 5. ОНВ, заканчивающиеся на гласную букву «й». При создании аориста, буква «й» в конце заменяется на «и». Личные окончания применяются по принципу типа F 2, т.е. местоименные суффиксы из 2-го вида (+ям, +й, +яд, +ем, +ед, +янд).
- Примеры: *зистан* (жить), ОНВ – *зй*, аорист – *зйям* (живу), *зистй* (живешь) и т.д.
- Для создания каузативов двух степеней используются конструкции ОНВ+ён(и)дан, ОНВ+ёнон(и)дан соответственно: *зиёндан* (заставить жить), *зиёнондан* (заставить жить через кого-то).
6. F 6. Недостаточные (безличные) глаголы
- 6.1. F 6.1. Глагол *боистан*. Этот глагол не спрягается, есть всего несколько форм, такие как *бояд*, *набояд*, *мабояд*, *мебоист* и т.д.
- 6.2. F 6.2. Глагол *шоистан*. Этот глагол тоже не спрягается, есть только 1 форма *шояд*.
7. F 7. Глаголы, имеющие 2 вида ОНВ и общий ОПВ. Словоформы образуются каждого вида ОНВ отдельно и личные окончания, применяются по принципу типа F 1.
- Примеры: *додан* (дать), ОНВ – *деҳ*, *диҳ*, аорист – *деҳам*, *диҳам* (отдаю), *диҳй*, *деҳй* (отдашь) и т.д.; *ниҳодан* (ставить), ОНВ – *неҳ*, *ниҳ*, аорист – *неҳам*, *ниҳам* (ставлю), *неҳй*, *ниҳй* (ставишь) и т.д.
- Для создания каузативов двух степеней используются конструкции ОНВ+он(и)дан, ОНВ+онон(и)дан соответственно: *деҳондан* или *диҳондан* (заставить отдать), *деҳонондан* или *диҳонондан* (заставить отдать через кого-то).
8. F 8. Глаголы, имеющие 2 вида ОНВ и общий ОПВ. При добавлении префикса *би*, буква «о» вначале основы заменяется на «ё».
- 8.1. F 8.1. Глагол *омадан* (прийти). При создании аориста, личные окончания применяются по принципу типа F 2, т.е. местоименные суффиксы из 2-го вида (+ям, +й, +яд, +ем, +ед, +янд). ОНВ – *о*, *ой*, формы с префиксом – *биёям* (приду), *биёй* (придешь), *биё* (приди), *биёед* (придите) и т.д.
- Для создания каузативов двух степеней используются конструкции *би+ОНВ+ён(и)дан*, *би+ОНВ+ёнон(и)дан* соответственно: *биёёндан* (заставить прийти), *биёёнондан* (заставить прийти через кого-то).
- 8.2. F 8.2. Глагол *овардан* (принести). При создании аориста, личные окончания, применяются по принципу типа F 1 (+ам, +й, +ад, +ем, +ед, +анд). ОНВ – *ор*, *овар*, формы с префиксом – *биёрам*, *биёварам* (принесу), *биёр*, *биёвар* (принеси) и т.д.
- Для создания каузативов двух степеней используются конструкции *би+ОНВ+он(и)дан*, *би+ОНВ+онон(и)дан* соответственно: *биёрондан* (заставить принести), *биёронондан* (заставить принести через кого-то).
9. F 9. Вспомогательный глагол *ҳаст* (не быть) и его отрицание *нест* (не быть, нет). Есть всего несколько форм, а именно спряжение по лицам и числам: *ҳастам* (я есть), *ҳастй* (ты есть), *ҳаст* (он есть) и т.д.; *нестам* (я не есть, меня нет), *нестй* (ты не есть, тебя нет), *нест* (он не есть, его нет) и т.д. Нет каузативных форм.
- В результате классификаций, 181 глаголов были распределены по типам формообразования как показано в табл. 12.

Таблица 12. Распределение по типам формообразований глаголов

Тип формообразования	Количество соответствующих типу слов	Количество словоформ одного слова
F 1	147	1503
F 2	13	1518
F 3	2	1518
F 4	10	1503
F 5	1	1518
F 6.1	1	8
F 6.2	1	1
F 7	2	2655
F 8.1	1	1518
F 8.2	1	2655
F 9	2	12
Всего	181	

ХIII. ГЕНЕРАЦИЯ ТАДЖИКСКИХ СЛОВОФОРМ ГЛАГОЛА В ИНТЕРНЕТ-ПРИЛОЖЕНИИ

Программная реализация классификаций глагола, а также алгоритм генерации форм слов, реализована в Интернет-приложении «Ибора» (тадж. «словосочетание»), расположенная по адресу <https://ibora.tk> (рис. 1). Данная реализация является частью приложения, в котором ранее были выполнены классификации существительных, прилагательных и местоимений. При разработке этого Интернет-приложения использовались язык программирования PHP7 и фреймворк CodeIgniter 3.1.10. Словарь Интернет-приложения и базы постфиксов могут расширены по мере необходимости.

Ибора

← кушодан

кушоёнад, бикишоёнад, букушоёнад, макушоёнад, мекушоёнад, накушоёнад, бинакушоёнад, мебикушоёнад, менакушоёнад, набикушоёнад, нобикушоёнад, ҳамекушоёнад, ҳамебикушоёнад, ҳаменакушоёнад, намекушоёнад, кушояд, бикушояд, букушояд, макушояд, мекушояд, накушояд, бинакушояд, мебикушояд, менакушояд, набикушояд, нобикушояд, ҳамекушояд, ҳамебикушояд, ҳаменакушояд, намекушояд, кушоёнам, бикишоёнам, букушоёнам, макушоёнам, мекушоёнам, накушоёнам, бинакушоёнам, мебикушоёнам, менакушоёнам, набикушоёнам, нобикушоёнам, ҳамекушоёнам, ҳамебикушоёнам, ҳаменакушоёнам, намекушоёнам, кушодам, кушоён(и)дам, кушоён(и)дам, кушоён(и)даму, макушоён(и)дам, намекушоёнанду, кушоянду, бикушоянду, букушоянду, макушоянду, мекушоянду, накушоянду, бинакушоянду, мебикушоянду, менакушоянду, набикушоянду, нобикушоянду, ҳамекушоянду, ҳамебикушоянду, ҳаменакушоянду, намекушоянду, кушову, бикушову, букушову, макуш...

Количество словоформ – 1518

[К списку глаголов](#)

Проект «Ибора» © 2018-2022 Навруз Мадибрагимов
Сайт находится в стадии тестирования и отладки

Рис.1. Страница сгенерированных словоформ глагола «кушодан» (русс. «открыть»)

XIV. ЗАКЛЮЧЕНИЕ.

В процессе исследования морфологии таджикского языка были получены следующие результаты.

1. Классифицировано 181 глаголов таджикского языка.
2. Определены 9 типов и их 2 подтипа формообразования слов глаголов. Для каждого типа и подтипа выделены характерные особенности.

Классификация глаголов таджикского языка по типам формообразования используется для генерации их словоформ в разработанном Интернет-приложении. Полученные результаты генерации словоформ были проверены и одобрены специалистами по морфологии таджикского языка.

В настоящее время продолжается исследование таджикской морфологии. Исследуются и классифицируются такие части речи как, числительное, наречие, при-

частие и деепричастие. Результаты классификаций в будущем будут опубликованы. Интернет-приложение «Ибора», после доработок должна служить инструментом для автоматической обработки таджикских текстов, что является полезным для исследователей и интересующимся этой сферой науки.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- [1] Мадибрагимов Н.Ш., Пруцков А.В. Классификация существительных таджикского языка для автоматической обработки текстов // Прикаспийский журнал: управление и высокие технологии. – 2020. – № 4. – С. 39–52.
- [2] Мадибрагимов Н.Ш., Пруцков А.В. Типы прилагательных и местоимений таджикского языка и их использование для генерации словоформ // International Journal of Open Information Technologies. – 2021. – Т. 9. – № 11. – С. 85-89.
- [3] Пруцков А.В. Алгебраическое представление модели формообразования естественных языков // Cloud of Science. – 2014. – Т. 1. – № 1. – С. 88-97.
- [4] Ниязмухаммадов Б., Бузург-зода Л. Морфология таджикского языка. На таджикском языке. – Сталинабад: Таджикгосиздат, 1941. – 70 с.
- [5] Саймиддинов Д., Холматова С.Д., Каримов С. Таджикско-русский словарь. – Душанбе: Пайванд, 2006. – 784 с.
- [6] Арзуманов С.Д., Джалолов О.Д. Учебник таджикского языка для вузов. – Душанбе: Ирфон, 1969 – 415 с.
- [7] Иванов В.Б., Семёнова Е.В., Хушкадамова Х.О. Учебник таджикского языка для стран СНГ: в 2 ч. Ч. 1 / Мос. гос. ун-т им. М.В. Ломоносова, Ин-т стран Азии и Африки. – М.: Ключ-С, 2009. – 232 с.
- [8] Арзуманов С.Д., Сангинов А. Таджикский язык. – Душанбе: Маориф, 1988. – 416 с.
- [9] Довудов Г.М. Компьютерный морфологический анализ таджикских словоформ: дис... канд. техн. наук: 05.13.11; защищена 06.04.18. – Душанбе, 2018. – 161 с.

Features of Computer Morphological Analysis and Synthesis of verbs of the Tajik language

Navruz Madibragimov

Abstract — Today, computational linguistics of the Tajik language is at the origin of its development. In order to develop this area, the author of this article is developing a project for the formalization of inflections of the Tajik language for computer morphological analysis and synthesis, within which it is required to classify Tajik words according to the types of formation. The words nouns, adjectives and pronouns are already has been classified. In this article, the author proposes a classification of the verbs of the Tajik language. The article highlights 9 types of verb formation and 4 subtypes. For the presented types and subtypes, their distinctive features are shown. This classification is a continuation of the author's research, which includes the classification of nouns, adjectives and pronouns in the Tajik language. The author has developed a program as an Internet application to implement the generation of word forms in the Tajik language. This application is powered by a linguistic base completed as a result of ongoing research in this article. Classification of other parts of speech of the Tajik language continues.

Keywords—computational linguistics, natural language processing, Tajik language, morphology of the Tajik language, morphogenesis model, generation of word forms, determining word forms, Internet application.

REFERENCES

- [1] Madibragimov N.Sh., Prutzkow A.V. Klassifikatsiya sushchestvitel'nykh tadjhikskogo yazyka dlya avtomaticheskoy obrabotki tekstov [Classification of nouns of the Tajik language for natural language processing] // Caspian journal: management and high technologies, 2020, no 4, pp: 39–52.
- [2] Madibragimov N.Sh., Prutzkow A.V. Tipy prilagatel'nykh i mestoimennykh tadjhikskogo yazyka i ikh ispol'zovaniye dlya generatsii slovoform [Types of adjectives and pronouns of the Tajik language and their use to generate and determining word forms] // International Journal of Open Information Technologies, vol. 9, no. 11, 2021.
- [3] Prutzkow A.V. Algebraicheskoye predstavleniye modeli formoobrazovaniya yestestvennykh yazykov [Algebraic representation of the morphogenesis model of natural languages] Cloud Of Science. 2014, vol. 1, no 1, pp. 88-97.
- [4] Niyazmukhammadov B., Buzurg-zoda L., Morfologiya tadjhikskogo yazyka. Na tadjhikskom yazyke [Morphology of the Tajik language. In Tajik language]. Stalinabad: Tadjhikgosizdat, 1941.
- [5] Saymiddinova D., Kholmatova S.D., Karimov S. Tadjhiksko-russkiy slovar' [Tajik-Russian Dictionary]. Dushanbe: Payvand, 2006.
- [6] Arzumanov S.D., Jalolov O.D., Uchebnik tadjhikskogo yazyka dlya vuzov [Tajik language textbook for universities]. Dushanbe: Irfon, 1969.
- [7] Ivanov V.B., Semyonova E.V., Khushkadamova Kh.O. Textbook of the Tajik language for countries: in 2 parts, Part 1 / Moscow State University named after M.V. Lomonosov, Institute of Asian and African countries. M.: Klyuch-S, 2009. 232 p.
- [8] Arzumanov S.D., Sanginov, A. Tadjhikskiy yazyk [Tajik language]. Dushanbe: Maorif, 1988. 416 p.
- [9] Dovudov G.M. Komp'yuternyy morfologicheskiy analiz tadjhikskikh slovoform [Computer morphological analysis of Tajik word forms]. Dushanbe, 2018. 161 p.

[1] Madibragimov N.Sh., Prutzkow A.V. Klassifikatsiya sushchestvitel'nykh tadjhikskogo yazyka dlya avtomaticheskoy