

## Фреймовый подход при генерации форм персидских глаголов

Луканин Артём Викторович  
Южно-Уральский государственный университет

Современный персидский язык относится к иранской группе индоевропейской семьи языков и имеет богатую словоизменительную парадигму. Черты флективного строя сохранились в глаголе, где личные окончания совмещают в себе значения лица и числа. Наряду с этим имеются многочисленные аналитические формы глагола, состоящие из причастия или (реже) усечённой формы инфинитива и личных форм вспомогательных глаголов. В связи с трудностью обучения спряжению персидских глаголов создана программа автоматической генерации парадигм глагола Persian Verb Conjugator (PVC), основанная на фреймовом подходе. Система PVC расположена по адресу: <http://staging.laits.utexas.edu/persian/pvc/>.

### 1 Описание системы

В PVC каждой форме глагола соответствует строго определённый фрейм, слоты которого либо остаются пустыми, либо заполняются псевдо-аффиксами, вычисляемыми для каждого глагола на основе фонологических правил. Под псевдо-аффиксом понимается последовательность букв, неизоморфная суффиксу в традиционной грамматике. К примеру, инфинитив «sukhtan» (سوختن) делится на первичное причастие (the primordial participle) «sukh» и суффикс «tan» в Академической грамматике современного персидского языка<sup>1</sup>, в то время как в PVC инфинитив делится на основу прошедшего времени «sukht» и псевдо-аффикс «an».

PVC автоматически генерирует все словоформы для следующих видо-временных форм, включая положительные и отрицательные формы:

1. Простое прошедшее время (Simple Past).
2. Прошедшее длительное время (Imperfect).
3. Прошедшее время в сослагательном наклонении (Perfect Subjunctive).
4. Прошедшее определённое время (Past Progressive).
5. Перфект (основная форма) (Present Perfect).
6. Преждепрошедшее время (Past Perfect).
7. Настояще-будущее время (Present Indicative).
8. Настоящее определённое время (Present Progressive).
9. Настояще-будущее время (аорист) в сослагательном наклонении (Present Subjunctive).
10. Будущее категорическое время (Future).

## 11. Повелительное наклонение (Imperative).

Порядок следования и английские названия видо-временных форм соответствуют порядку и названиям в системе PVC, созданной в помощь студентам Университета Техаса в Остине (США). Русские названия видо-временных форм соответствуют терминологии В.В. Мошкало<sup>2</sup>. Длительный перфект и преждепрошедший перфект нами не рассматривался. В персидском языке нет отрицательных форм в настоящем определённом и прошедшем определённом временах. Для целей обучения было решено также исключить из парадигмы повелительного наклонения формы 1 и 3 лица. Таким образом, для большинства глаголов в PVC выводится 112 словоформ. Тем не менее, функция генерации форм глагола, используемая как в системе PVC, так и в тестах по спряжению, может сгенерировать 240 словоформ (120 в активном залоге и 120 в пассивном залоге). В действительности в функции генерации содержится 120 фреймов для генерации форм глагола в активном залоге и 1 фрейм для генерации причастия прошедшего времени, который в соединении с фреймами активного залога формирует все формы пассивного залога.

В строку ввода PVC можно вводить инфинитив глагола либо буквами персидского алфавита, либо в транскрипции. Перед генерацией словоформ определяется алфавит, с помощью которого записан глагол, и выводятся все парадигмы в соответствующем алфавите.

Для записи глаголов не требуется дополнительных средств ввода букв. Страница PVC оснащена визуальной клавиатурой персидского языка, которая позволяет вводить персидские буквы в кодировке Unicode (UTF-8). Используемая транскрипция персидского языка содержит только 1 букву, отсутствующую на большинстве клавиатур – «ā». Вместо неё можно вводить последовательность «aa» (обычная практика носителей персидского языка в повседневной деятельности, к примеру, при электронной переписке), которая автоматически преобразуется в «ā».

## 2 Две основы глагола

Простых глаголов в персидском языке сравнительно мало. Большинство глаголов образуется из простых с помощью приставок и суффиксов. Также широко используются так называемые сложные (составные) глаголы, которые образуются сочетанием имени и глагола<sup>2</sup>. Каждая словоформа персидского глагола образуется на базе одной из двух основ: основы настоящего времени и основы прошедшего времени. Основа прошедшего времени автоматически получается из инфинитива. Трудность представляет получение основы настоящего времени. Карин Мегердумизн отмечает, что данные основы невыводимы одна из другой<sup>3</sup>, однако, Джон Эндрю Бойл предложил 10 групп глаголов, имеющих похожие соответствия между основой настоящего и основой прошедшего

времен<sup>4</sup>. Взяв за основу эти 10 групп, нами было создано 10 правил вывода основы настоящего времени из основы прошедшего времени. Данные правила позволяют получить корректные основы настоящего времени для многих составных глаголов и глаголов, образованных с помощью префиксального словообразования, т.к. для выбора правила используется псевдо-окончание глагола. Исключения из правил записаны в лексиконе исключений. Составные глаголы, имеющие в составе глаголы-исключения, также интерпретируются системой корректно, однако, т.к. в системе PVC нет особого правила для получения леммы глагола-исключения из глагола, образованного из него с помощью префиксального словообразования, то каждый глагол-исключение должен быть записан в лексиконе вместе со всеми его производными.

### 3 Фреймы словоформ глагола

Каждая словоформа в системе PVC описана в виде фрейма с определённым набором слотов. Слоты заполняются псевдо-аффиксами, предварительно вычисленными для каждого глагола. Псевдо-аффикс может отсутствовать, а значит и слот, соответствующий ему, будет пустым. К примеру, слот COMPOUND заполняется именем только тогда, когда введённый глагол является составным, образованным с помощью имени. Аналогично, слот VyV, соответствующий межвокальной йотированной согласной (персидскому йе), будет пустым для глагола ‘bar gozidan’ (برگزیدن) и заполнен для глагола ‘afzudan’ (افزودن) в настоящем определённом времени в 1 лице ед. числа, имеющем форму ‘dāram mīafzāyam’ (دارم می‌افزایم).

Система PVC и тесты, основанные на ней, прошли апробацию и используется в учебном плане Университета Техаса в Остине, Университета Мичигана и колледжа The South Orange County Community College District.

---

<sup>1</sup> Navid Fazel. 2006. *Academic Grammar of New Persian*. <http://www.fazel.de/dastur/EN/index.html>

<sup>2</sup> Мошкало В.В. Персидский язык // Языки мира / РАН. Ин-т языкознания; Редкол.: В.Н. Ярцева (пред.) и др. - М.: Индрик, 1997. - Иранские языки. Ч. 1: Юго-западные иранские языки / Редкол.: В.С. Расторгуева (отв. ред.) и др. - С. 71-102.

<sup>3</sup> Karine Megerdomian. *Finite-state morphological analysis of Persian*. In Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages. Coling 2004, University of Geneva. August 28, 2004.

<sup>4</sup> Boyle, John Andrew. *Grammar of modern Persian*. Wiesbaden, Harrassowitz, 1966.