

SLPO: A Consistent Persian Orthography Based on the Latin Script

Reza Salarmehr
Amirkabir University
salarmehr@aut.ac.ir

Abstract

Persian is officially written in Arabic script which causes some well-known problems in writing academic material for native and non-native speakers. In this paper we propose a standard romanization and orthography scheme for Persian named *SLPO* that is best fitted for natural language processing, learning purposes and writing academic material (but not limited to) and we claim that it is the most comprehensive and consistent scheme ever presented for this language.

Keywords: romanization scheme, Persian, natural language processing, orthography

1 Introduction

In Iran and Afghanistan, Persian is officially written in Arabic script that is a right to left script (RTL). In this script short vowels are usually not written, and as a result a lot of word-forms have more than one pronunciation and meaning (e.g: (سیر، بر، کرم). This fact makes writing and reading Persian hard for non-native (and even native) speakers [?]. As a result most authors prefer to use a romanized form of Persian words that is easier to read and is supported better in computer environments.

However, the absence of an accepted romanization standard caused authors to use various personal romanization schemes that in turn caused some problems. In this paper, we represent two romanization schemes for two distinct use-cases. We claim that they are the most suitable schemes available to be used in general and particularly in academic cases. The first scheme is intended to be used for general applications but it is not totally reversible, and the second one is intended to be used in situations that reversibility is important. We have provided a web site that contains the Latin form of more than 350,000 Persian words and idioms. You can access it in <http://vajje.com>.

The structure of this paper is as follows: in the next section we draw an overview of related works. In section ?? we point to characteristics of our proposed orthography named SLPO. In sections ?? and ?? romanization and orthography issues are discussed respectively.

2 The Arabic Script Problems

The problems of the Arabic script for writing Persian have been investigated extensively during the two last centuries [?]. A complete review of this topic is out of the scope of this paper. Here, we point to two main enigma regarding using the Arabic script to write Persian (For a comprehensive overview of related literature refer to [?]).

2.1 Short Vowel Enigma

Short vowels (/æ/, /e/, /o/) are not generally represented in written Persian while changing the short vowel in a word can create a totally unrelated new word. As an example <گل> represents two unrelated words: <gel> (*mud*) and <gol> (*flower*). So correct pronunciation of words which contains a short vowel is not possible without prior knowledge and if a word-form represents more than one word, the reader should choose correct pronunciation according to its context. As a real example see ?? figure. The name of a single alley could not be pronounced correctly by the mayoralty staffer and is romanized differently on two signboards in the alley.

Ezafé is a one of most important grammatical features of Persian. Ezafé is pronounced as a short vowel, /e/, and similar to other short vowels it is generally not written. If the word preceding Ezafé ends in a vowel, Ezafé will be articulated /je/ and it will have a written presentation, <ی>. Also there is no consensus on how to write the Persian indefinite article when the referent ends in /e/. As an example <ی>+<خانه> may be



Figure 1: Two signboards in Tehran. The name of the alley, <پشن>, could not be pronounced correctly and is romanized in two different forms in the same alley.

wrote as <خانه‌ای>, <خانه‌یی> or <خانه‌ئی> [?].

2.2 ZWNJ Enigma

Appearance of Arabic letters changes depending on the letters that precede or succeed them in a word and the structure of the word that the letter belongs to. For example, the letter sin (<س>) has these forms: <س>, <س>, <س> and <سس>. When writing Persian compound words, some times letters in the concatenation point join each other (e.g. <گل>+<آب>→<گلاب>) and sometimes they not (e.g. <جسته>+ZWNJ+<گریخته>→<جسته‌گریخته>). In computer typography a zero width non-joiner character (ZWNJ) is inserted between stems to prevent them from joining each other. Unfortunately, there is no accepted rule for writing derived and compound words (e.g. <duncjoo> is written as <دانشجو> and <دانش‌جو> [?, ?, ?, ?]).

Similar to other Semitic languages, Arabic has a nonconcatenative "root-and-pattern" morphology. The pattern can be used to read words correctly even without presenting short vowels. Compounding and derivation are not used in the Arabic language to form new words. As a result, the above two enigmas are not relevant in that language. Indeed the Arabic script only works well for writing the Arabic language.

2.3 Why is a standard romanization scheme necessary?

In addition to the above issues, there are some other difficulties that caused those whose first language is not Persian to use the Latin script, instead of the Arabic script during their learning course or in their publications. Unfortunately, each author uses her/his personal romanization scheme that makes reading and indexing of romanized Persian texts difficult. Even in Iran when there is a need to romanize the name of people and places, there is no unique accepted romanization scheme. Consequently different romanized forms may be used by distinct users which make searching the name of people and places hard and confusing.

3 Related works

In this section, we overview Persian romanization schemes that have been proposed in the past twenty years.

Maleki [?] proposed a romanization scheme which contains 29 phonemes (23 consonants and 6 vowels) named *Dabire* (formerly *eFarsi* [?]). *Dabire* uses a single Latin grapheme for Arabic graphemes that have the same phonology in Persian. The author provides some guidelines for writing compound words, abbreviations and foreign load words. He also describes capitalization and punctuation in his romanization scheme.

Ghayour [?] provides a romanization mapping table named *Ironic*. Using the letter <w> to represent /u:/, his scheme only includes basic Latin alphabet. In his scheme letters <a>, <e>, <o>, <i>, <u> and <w> respectively represent /æ/, /e/, /v:/, /i:/, /o/ and /u:/. Unfortunately, this caused the well-known pronunciation of letters <o>, <u>, <j> and <w> to be changed. Currently, Iranians generally pronounce these letters /o/, /u:/, /dʒ/ and /v/, respectively. Furthermore it uses a digraph for /dʒ/ and a monograph for /z/ while /z/ has the least frequency in Persian.

Moslehi [?] introduced a romanization scheme named *IPA2 (Pársik)*. He uses a lot of diacritics to represent ayn and hamza (called *mul* in IPA2). He also used a digraph for /ʃ/ and a monograph for /tʃ/, while <ش> is a frequent Persian letter and <چ> has very low frequency in Persian. The letter <w> is also used in some words, e.g. <موز> = <moz>. The scheme also has some rules for writing in informal language and some other guidelines.

Mahdavi [?] proposed a strict mapping from Arabic script to Latin script that maintains all Arabic diacritics and even radical and ambiguous forms of letters (e.g. <ف>). The scheme contains almost 80 characters. Actually, adding Arabic diacritics to Persian texts is simpler to learn and use. Nevertheless his suggestion may be useful in processing the old Persian and Arabic manuscripts.

There are also some other rudimentary schemes, including: The Tajik alphabet in Latin (obsolete), Paarsi [?], UniPers [?], UN romanization of Persian for Geographical Names [?], Library of Congress/American Library Association romanization of Persian, IJMES transliteration system for Arabic, Persian, and Turkish [?], Encyclopaedia Iranica's scheme [?].

All proposed schemes suffer from at least two of the following deficiencies: (i) they use some new or modified Latin letters that make their use hard for Persian-learners. (ii) they ignore phonological rules of Persian and spell some words unnecessarily long (iii) they lack a firm rule for writing saken ayin and hamza. (iii) they do not comprehensively describe orthography of all Persian grammatical structures. (iv) they do

not suggest a romanization scheme in situations in which reversibility matters. (v) they ignore the semantic accepts of the writing system and do not provide any clues to shed light on the meaning of words (vi) they do not provide a firm regular method to spell Persian verbs.

In this paper we have proposed a new orthography scheme which addresses all of these issues.

4 SLPO

In this section we introduce our suggested romanization schema named *Standard Latin-based Persian orthography* abbreviated as SLPO.

In developing our romanization scheme we used some statistics data such as letter frequency and phoneme frequency. You can see Persian letter frequency in figure ?? and ??.

SLPO is the most comprehensive and consistent Persian romanization scheme ever proposed. Some of its applications include:

- Writing academic materials such as papers and books.
- Teaching Persian to non-native learners.
- Transliteration of geographical names
- Natural language processing and text mining

4.1 Priorities

The orthography is based on the following important priorities.

1. Unequivocalness and Reversibility: The words and sentences should be unequivocal as much as possible. This means that a single word-form should represent as few words as possible.
2. Conciseness: The least amount of glyphs should be used for writing words and phrases.
3. Phonetic: The orthography should be highly regular and word's pronunciations could be identified by their spelling.
4. No new glyph: SLPO uses only basic Latin alphabets. As a result it can be entered in all computational environments without any modifications.
5. Easy to learn, easy to use: The proposed scheme is easy to learn and use even for new learners.

5 General Romanization

The Latin script of Persian consists of 25 letters. It does not use the letter <w>. You can see the letters in table ?. There are some rules in using this alphabet as follow:

Rule 1 *Each word contains at least one vowel letter (<a,e,i,o,u>).*

Rule 2 *No word begins with more than two consonants.*

See section ?? to see how to pronounce words that begin with two consonants.

5.1 Vowels

Persian has six vowels, including: /ɒ:/, /e/, /o/,/æ/, /i:/ and /u:/. The first three rows of table ?? show how a single word-form in the Arabic script can be pronounced differently and yet convey distinct semantics. In the following subsection we explain how to map each vowel to its equivalent in Latin script.

Table 2: Persian vowels

Latin	IPA	Arabic
sar	/sæɾ/	سر
ser	/seɾ/	سر
sor	/soɾ/	سر
sur	/sɒ:r/	سار
soor	/su:r/	سور
sir	/si:r/	سیر

5.1.1 A

Letter A represents the *near-open front unrounded vowel* (/æ/). Table ?? shows some examples of its usage.

Table 3: Examples for A

Latin	Arabic
adab	ادب
moallem	معلم
moarrefi	معرفی
akkus	عکاس
Aruk	اراک
Alueddin	علاءالدین

Table 1: Letters of general scheme

Latin Letter	IPA	Arabic counterparts	Cyrillic counterpart(s)
A a	/æ/	أَ، عَ، ئَ	А а, (Ѡ)
B b	/b/	ب	Б б
C c	/ʃ/	ش	Ш ш
D d	/d/	د	Д д
E e	/e/	إِ، يَ، عِ	Е е, Э э
F f	/f/	ف	Ф ф
G g	/g/	گ	Г г
H h	/h/	هَ، حَ	Һ һ
I i	/i:/	يَ؛ اِي، ئِي، عِي، يِي	И и, Й й, (Ё ё)
J j	/dʒ/	ج	Й й
K k	/k/	ك	К к
L l	/l/	ل	Л л
M m	/m/	م	М м
N n	/n/ /ŋ/	ن	Н н
O o	/o/	أُ، عُ، ئُ	У у
P p	/p/	پ	П п
Q q	/g/	ق، غ	Қ қ, Ғ ғ
R r	/r/	ر	Р р
S s	/s/	س، ث، ص	С с
T t	/t/	ت، ط	Т т
U u	/u:/	أُ، آ، ئَا، عَا	О о
V v	/v/	و	В в
X x	/x/	خ	Х х
Y y	/j/	ي	Й й
Z z	/z/	ز، ذ، ض، ظ	З з
Digraphs			
oo	/u:/	او، ئو، عو، وُ	Ӯ ӯ, (Ю ю, Я я)
jj	/ʒ/	ژ	Ж ж
cc	/tʃ/	چ	Ч ч

Rule 3 *When the letter <a> is placed at the end of a word or when it is preceded by a vowel, it represents glottal stop(/ʔ/) and so in these cases it is a consonant*¹

In such situations, the corresponding Arabic letter is *saken ayin*² (ع) or *saken hamza* (ء). Table ?? shows some examples of this usage (the words <jame> and <jamee> in this tables are only for comparison and do not contain consonant <a>).

Table 4: Examples for consonant A

Latin	Arabic
boad	بُعد
baad	بَعد
cear	شعر
roab	رعب
Saadi	سعدى
roayu	رؤيا
naal	نعل
Kaabe	كعبه
raay	رأى
nafa	نفع
nufea	نافع
jume	جامه
jumea	جامع
jumee	جامعه
joz	جز
joza	جزء
mzua ³	امضاء
ncua	انشاء

Rule 4 *If /a/ phoneme is preceded by another vowel it should be represented using two adjacent letters <a>.*

Look at table ?? to see some examples of this rule.

Table 5: Examples for <aa>

Latin	Arabic
coaab	شُعَب
coaartua	شعرا
zoaafu	ضعفا
biaadab	بيادب

¹The only exception for this rule is the word <na> (=no in English) that is pronounced: /næ/.

²A consonant that is not followed by a vowel is a *saken* consonant.

³Section ?? explains how to use two consonants in the beginning of words.

5.1.2 E

Letter <E> presents the *close-mid front unrounded vowel* (/ɛ:/). Table ?? shows some examples of its usage.

Table 6: Examples for E

Latin	Arabic
enuyat	عنایت
nume	نامه
enekus	انعكاس
peste	پسته
Eruq	عراق
eatemud	اعتماد
erue	ارائه
majmooe	مجموعه

5.1.3 I

The letter <i> represents the *long lose front unrounded vowel* (/i:/). Table ?? shows some examples of its usages.⁴

Table 7: Examples for I

Latin	Arabic
sib	سبب
Irun	ایران
imun	ایمان
Said	سعید
ide	ایده
id	عید
Nuin	نابین
niuz	نیاز

Notice that unlike English, in Persian two adjacent vowels do not form a digraph and they should be pronounced separately (except <oo> digraph). Table ?? shows some examples of adjacent vowels.

The pronunciation of <i> when is succeeded by a vowel is comparable with the letter <i> in English.

⁴In all tables *Arabic* refers to the Arabic script (Perso-Arabic script) and *Latin* refers to the Latin script

Table 8: Examples for I and an adjacent vowel

Latin	Arabic	
a+i	rais	رئيس
e+i	perotein	پروتین
o+i	moin	معین
oo+i	dastcooi	دست شویی
i+i	tabii	طبیعی
i+a	tacia	تشیع
i+e	cie	شیعه
i+o	biorze	بی عرضه
i+a	piuz	پیاز
i+oo	Nioocu	نیوشا
i+i	cii	شیعی

5.1.4 O

The letter <o> represents the *close-mid back rounded vowel* (/o/). Table ?? shows some examples of its usage.

Table 9: Examples for O

Latin	Arabic
ordak	اردک
tond	تند
bord an	بُردن
omid	امید
Oroopu	اروپا
Oroomie	ارومیه
Osetestun	استستان

5.1.5 U

The letter <u> represents the *long open back rounded vowel*, (/u:/). Table ?? shows some examples of its usage.

Table 10: Examples for A

Latin	Arabic
ub	آب
ulem	عالم
ulam	عالم
reuyat	رعایت
haruyene	هرآینه
uyene	آینه
Umol	آمل
Ubxuzestun	آبخازستان
Qorun	قرآن

5.1.6 The digraph oo

The digraph <oo> is pronounced /u:/ (long close back rounded vowel). Table ?? shows some examples of its usage.

Table 11: Examples for digraph <oo>

Latin	Arabic
rooz	روز
poost	پوست
holoo	هلو
moosiqi	موسیقی
noocid an	نوشیدن
raoof	رئوف
masool	مستول
oo	او
ood	عود
hayoolu	هیولا
yooz	یوز
Yoosof	یوسف
Yoonun	یونان

5.2 Consonants

The Latin alphabet of Persian is consists of 20 consonant letters, as described below. Most of the examples that follow are intended to be used as a pattern for writing other Persian words.

5.2.1 B

The letter is always used to represent the *voiced bilabial stop* (/b/) (like most other languages that use Latin Scripts). Its counterpart in Arabic script is <ب>. Table ?? shows some examples of its usage.

Table 12: Examples for B

Latin	Arabic
babr	ببر
morabbu	مرّبا
bomb	بمب
tarbiat	تربیت
Birjand	بیرجند

5.2.2 C

When this letter is not used in the digraph <cc> it is always used to represent the *voiceless palato-alveolar fricative* (/ʃ/). (Similar to its usage in Zhuang and Kabyle alphabets.) Its

counterpart in Arabic script is <ش>. Table ?? shows some examples of its usage.

Table 13: Examples for C

Latin	Arabic
cir	شیر
cokolut	شکلات
tacakkor	تشکر
ncuallah	انشاءالله
cirin	شیرین
Cirin	شیرین
Cuc	شوش

5.2.3 D

The letter D always represents the *voiced alveolar plosive* (/d/), (like nearly all languages using the Latin script). Its counterpart in Arabic script is <د>. Some examples of its usage are shown in table ??.

Table 14: Examples for D

Latin	Arabic
dooq	دوغ
dur	دار
dar	در
taraddod	تردد
moaaddab	مؤدب
Docanbe	دوشنبه
docanbe	دوشنبه

5.2.4 F

The letter <f> always represents the *voiced alveolar plosive* (/f/), (like most languages that use the Latin script). Its counterpart in Arabic script is <ف>. Some examples of its usages are shown in table ??.

Table 15: Examples for F

Latin	Arabic
fardu	فردا
Furubi	فارابی
tafuvot	تفاوت
tafakkor	تفکر
foroodguh	فرودگاه
foot	فوت

5.2.5 G

The letter <g> always represents *voiceless labiodental fricative* (/g/), (like a lot of languages that use the Latin script). Its counterpart in Arabic script is <گ>. Some examples of its usage are shown in table ??.

voiced velar stop

Table 16: Examples for G

Latin	Arabic
gol	گل
gel	گل
gelooguh	گلوگاه
sag	سگ
Gilun	گیلان
gooc	گوش
garm	گرم
geram	گرم

5.2.6 H

The letter <h> always represents the *voiceless glottal transition* (/h/), (like English, Faroese, German, Swedish etc. alphabets). Its counterparts in Arabic script are <ه> and <ح>. Some examples of its usage are shown in table ??.

Table 17: Examples for H

Latin	Arabic
hodhod	هدهد
hole	حوله
tahammol	تحمل
tahdid	تهدید
muh	ماه
neguh	نگاه
mehrub	محراب

5.2.7 J

When this letter is not used in the digraph <jz>, it always represents the *voiced palato-alveolar affricate* (/dʒ/),⁵ (like English, Portuguese and Turkmen alphabets). Its counterpart in Arabic script is <ج>. Some examples of its usage are shown in table ??.

⁵In this paper we use /dʒ/ to represent /dʒ̄/.

Table 18: Examples for J

Latin	Arabic
jun	جان
tajaddod	تجدد
Jukurtu	جاكارتا
jagure	جگاره
hejdah	هجده
joo	جو
jo	جو
jav	جو

5.2.8 K

The letter <k> always represents the *voiceless velar stop (/k/)* (like most languages that use the Latin script). Its counterpart in Arabic script is <ک>. Some examples of its usage are shown in table ??.

Table 19: Examples for K

Latin	Arabic
kooh	کوه
kuk	کاک
keyk	کيک
morakkab	مرکب
gerufik	گرافیک
kare	کره
korre	کره
kore	کره
korh	کره
dokme	دکمه
tekme	تکمه
Akbar	اکبر
Kubol	کابل
Alluhoakbar ⁶	الله اکبر

5.2.9 L

The letter <l> always represents the *alveolar lateral approximant (/l/)* (like most languages that use the Latin script). Its counterpart in Arabic script is <ل>. Some examples of its usage are shown in table ??.

⁶Common Arabic short sentences are written as a word in Persian

Table 20: Examples for L

Latin	Arabic
laqzid an	لغزیدن
lule	لاله
Lule	لاله
Leylu	لیلا
talaalo	تالالو

5.2.10 M

The letter <M> always represents the *bilabial nasal (/m/)* (like a lot of languages that use the Latin script). Its counterpart in Arabic script is <م>. Some examples of its usage are shown in table ??.

Table 21: Examples for M

Latin	Arabic
mur	مار
mudar	مادر
talammoz	تلمیز
moaasser	مؤثر
mozuik	موزائیک

5.2.11 N

The letter <n> when preceded by letters <K> or <G> represent *velar nasal (/ŋ/)* and in all other cases represents *alveolar nasal (/n/)* (like a lot of languages that use the Latin script). Its counterpart in Arabic script is <ن>. Some examples of its usage are shown in table ??.

Table 22: Examples for N

Latin	Arabic
nunvu	نانوا
Nijerie	نیجریه
moaannas	مؤنث
noxost	نخست

5.2.12 P

The letter <p> always represents *voiceless bilabial stop (/p/)* (like most of the languages that use the Latin script). Its counterpart in Arabic script is <پ>. Some examples of its usage are shown in table ??.

Table 23: Examples for P

Latin	Arabic
pu	پا
pool	پول
Pursi	پارسی
toop	توپ
pecc-pecc	پچ پچ
poocuk	پوشاک
poocak	پوشک

5.2.13 Q

The letter <q> always represents the *voiced velar fricative* (/q/) ⁷ Its counterpart in Arabic script is <ق>. Some examples of its usage are shown in table ??.

Table 24: Examples for Q

Latin	Arabic
qur	غار
kaluq	کلاغ
qaltun	غلطان
qucoq	قاشق
reqqat	رقت
Quen	قائن

5.2.14 R

The letter <r> always represents the *alveolar trill* (/r/) (like a lot of languages written in Latin script). Its counterpart in Arabic script is <ر>. Some examples of its usage are shown in table ??.

Table 25: Examples for R

Latin	Arabic
roostu	روستا
mehr	مهر
rob e anur	رَب انار
sulur	سالار
morabbu	مرتا
Rezu	رضا

5.2.15 S

The letter <s> always represents *voiceless alveolar sibilant* (/s/) (like almost all languages that use the Latin script). Its coun-

⁷According to dialects and words it also may be pronounced *voiceless uvular stop* /q/ or *voiced uvular stop* /ʁ/.

terparts in Arabic script are <س>, <ص> and <ث>. Some examples of its usage are shown in table ??.

Table 26: Examples for S

Latin	Arabic
asus	اساس
saboon	صابون
lase	لثه
seture	ستاره
staxr	استخر
sfenuj	اسفنج
stexure	استخاره
stebdud	استبداد
Sfand	اسفند

5.2.16 T

The letter <t> always represents *voiceless alveolar plosive* (/t/) (like the English, German and Norwegian languages). Its counterparts in Arabic script are <ت> and <ط>. Some examples of its usage are shown in table ??.

Table 27: Examples for T

Latin	Arabic
toot	توت
morattab	مرتب
tar	تر
tur	تار
tooti	طوطی
Oturod	عطارد
tarak	ترک
tark	ترک
tork	ترک
Tork	ترک

5.2.17 V

The letter <v> always represents the *voiced labiodental fricative* (/v/) (like a lot of languages that use the Latin script). Its counterpart in Arabic script is <و> (when it is used as a consonant). Some examples of its usage are shown in table ??.

Table 28: Examples for V

Latin	Arabic
vorood	ورود
vazaq	وزغ
nuv	ناو
Davood	داوود
navid	نوید

5.2.18 X

The letter <x> always represents *voiceless velar fricative* (/x/) (It's similar to Chi in the Greek script and Kha in the Cyrillic script). Its counterpart in Arabic script is <خ>. Some examples of its usage are shown in table ??.

Rule 5 The letter <v> after the letter <x> is not pronounced.

Notice this rule is important to satisfy the second priority. (see section ??)

Table 29: Examples for X

Latin	Arabic
Xodu	خدا
xerad	خَرَد
nax	نَخ
taxayyol	تَخْيَل
xiul	خيال
ceyx	شَيْخ
xerxere	خِرْخِرِه
xar	خَر
xur	خار
xvur	خوار
xun	خان
xvun	خوان
xust an	خاستن
xvust an	خواستن
xvubid an	خوابیدن
xorcid	خورشید
xord	خَرْد
xord	خورد
xvic	خویش
xic	خیش
Xvurazm	خوارزم
Xordud	خرداد

5.2.19 Y

The letter <y> always represents the *voiced palatal approximant* (/j/) (like English). Its counterpart in Arabic script is

<ی> (when it is used as a consonant). Some examples of its usage are shown in table ??.

Table 30: Examples for Y

Latin	Arabic
yus	یاس
yek	یک
Rey	ری
rey	ری
yur	یار
ccuy	چای
raiyyat	رعیت

5.2.20 Z

The Letter <z> always represents *voiced alveolar fricatives* (/z/) (like almost all languages that use the Latin script). Its counterparts in Arabic script are <ز>, <ذ>, <ض> and <ظ>. Some examples of its usage are shown in table ??.

Table 31: Examples for Z

Latin	Arabic
gozurdan	گزاردن
zanboor	زنبور
soozan	سوزن
arziz	ارزیز
Zohre	زهرة
zahre	زهرة
Zamin	زمین
zamin	زمین
Zanjan	زنجان
zorrat	ذرت
zolm	ظلم
tazud	تضاد
zoozanaqe	ذوزنقه

5.3 Digraphs

SLPO has two digraphs for presenting /ʒ/ and /tʃ/ both of which are phonemes with the least frequency in Persian.

5.3.1 Jj

The digraph <jj> represents the *voiced palato-alveolar fricative* (/ʒ/). Some examples of its usage are shown in table ??.

Table 32: Examples for <jj>

Latin	Arabic
pajjmord an	پژمردن
mojjde	مژده
perojje	پروژه
Pejmun	پژمان
jjimnustik	ژیمناستیک
Jjupon	ژاپن
Jjule	ژاله
Ajji	اژی
ajjdahu	اژدها

5.3.2 Cc

The digraph <cc> represents the *voiceless palato-alveolar affricate* (/tʃ/) ⁸. Some examples of its usage are shown in table ??.

Table 33: Examples for C

Latin	Arabic
ccup	چاپ
ccap	چپ
ccakuvak	چکاوک
moccule	مچاله
cceleccle	چلچله
ccacm	چشم

5.4 Other Rules

5.4.1 Two consonant in the beginning of words

Rule 6 *If a word begins with two consonants, the first consonant is pronounced by adding /e/ at its beginning and the second consonant is pronounced using the vowel succeeding it.*

Such words include a lot of Persian originated words (which begin with <اس>), a large number of Arabic loan words (in particular, words derived from these forms: افعال, استفعال and انفعال) and a considerable number of other Indo-European load words e.g. <اسکیت> (=skate). Examples of these words are presented in table ??.

Notice that rule ?? causes letter <e> to appear in the begging of some words e.g. <ecq> = <عشق>, <esm> = <اسم>, <enekus> = <انعکاس>, <emurat> = <امارت>.

5.4.2 Ayin

As is obvious in the previous sections, Arabic letters ayin and hazma are not written when they are not saken (ساکن). When

⁸In this paper, /tʃ/ have been used to represent /tʃ/.

Table 34: Two consonants in the beginning of words

Latin	Arabic
btekur	ابتکار
cqul	اشغال
dbur	ادبار
ftexur	افتخار
gzemu	اگزما
hsun	احسان
jbur	اجبار
ktesub	اكتساب
lzum	الزام
mcab	امشب
nfejur	انفجار
psilon	اپسیلون
qruc	اغراق
rtebut	ارتباط
staxr	استخر
ttehud	اتحاد
yvun	ایوان
xruj	اخراج
zterub	اضطراب

they are saken, they are always preceded by a vowel letter. In this situation, they are written as <a>.

5.4.3 Compound words

Persian has a lot of compound words. In Latin script, no space should be used between the building blocks of a compound word. Table ?? shows some examples of Persian compound words.

6 Reversible Romanization

If you want to map some texts in Arabic script to the Latin script and the reversibility of the result is your concern, you should use the *reversible alphabet* which adds a number of diacritics to the *general alphabets*. The new letters and diacritics are shown in table ?? ⁹

⁹Also if necessary you can use <␣> character (U+2423, OPEN BOX) to denote a space character that does not have a counterpart in Arabic text e.g. <ببخش> → <be␣baxc>.

Similarly you can denote a ZERO WIDTH NON-JOINER character in Arabic script using <◌> (U+02CC, MODIFIER LETTER LOW VERTICAL LINE) e.g. <تنها> → <tanha◌> and <تنها> → <tanha◌>.

¹⁰Actually only three in-use words have <ش>. They are: <ماء الشعير> = <muaoççair>, <بشاش> = <baççuc> and <عشاق> = <ocçuq>. So you seldom see <cç>.

Table 36: Reversible alphabet and their general alphabet counterparts

Latin	Arabic	note
xocbarxord	خوش پر خورد	
raftoumad	رفت و آمد	
costocoo	شستشو	eostecoo
didobuzdid	دید و باز دید	
goftogoo	گفتگو	goftegoo
rixtopuc	ریخت و پاش	
zadoband	زد و بند	
faruzonacib	فراز و نشیب	
Siosepol	سی و سه پیل	
Beytolmoqaddus	بیت المقدس	
Maxzanolasrur	مخزن السرار	
mottafeqolqol	متفق القول	
kisehavu	کیسه هوا	<i>airbag</i>
dastrasipaziri	دسترسی پذیری	<i>accessibility</i>
hezurpuyun	هزار پایان	
jastegorixte	جسته گریخته	
siuhsefid	سیاه سفید	
bolandqumat	بلند قامت	
qunoongoriz	قانون گریز	
hasticenusi	هستی شناسی	<i>ontology</i>
giuhcenusi	گیاه شناسی	
spanddoodkon	اسپند دودکن	
ubgarmkon	آب گرم کن	
budumhendi	بادام هندی	
pestecumi	پسته شامی	
mahallisuzi	محلّی سازی	<i>localization</i>
beynolmelalisuzi	بین المللی سازی	<i>internationalization</i>

Arabic	General	Reversible
عَ	a	a'
عِ	a	'a
ع	e	'e
عَءَ	o	'o
عَا	u	'u
عو	oo	'oo
عی	i	'i
ءَ	a	a'
ءِ	a	'a
ء	e	'e
ءَءَ	o	'o
ءَا	u	'u
ءو	oo	'oo
ئی	i	'i
ه	H h	H h
ح	H h	Ḥ ḥ
ت	T t	Ṭ ṭ
ط	T t	Ṭ̣̣ ṭ̣̣
ة	T t	Ṭ̣̣̣ ṭ̣̣̣
غ	Q q	Q̣ q̣
ق	Q q	Q̣ q̣
ز	Z z	Ẓ ẓ
ذ	Z z	Ẓ̣ ẓ̣
ض	Z z	Ẓ̣̣ ẓ̣̣
ظ	Z z	Ẓ̣̣̣ ẓ̣̣̣
س	S s	Ṣ ṣ
ص	S s	Ṣ̣ ṣ̣
ث	S s	Ṣ̣̣ ṣ̣̣
ش	cc	cc̣ ¹⁰
و	o	ọ
ال	(a/e/o)[l]	(ā/ē/ī)[l]

Table 37: Examples for the reversible alphabet

Arabic	Latin
امارت	`emurat
عمارت	`emurat
غذا	qazu
قضا	qazu
متوجه	motevajje
حيات	hayut
حياط	hayuṭ
تألم	ta'aallom
تعلم	tá aallom
معاصر	mó uşer
مآثر	ma'uşer
يُعد	boè d
يعد	baà d
رأى	raa'y
در	dorr
دور	dor
دور	door
خرد	xord
خورد	xord
روغن	roqan
آمدوشد	umadoçod
فوق العاده	foqōlu'ade
ابو الفضل	Abōlfazl
بالاخره	bēlaxare
بيت المال	beytōlmul
عبدالرحمان	'Abdōrrahmun
عبدالله	'Abdōlluh
دايرة المعارف	duyeraṭqolmá uref
متفق القول	mottafeqolqol
علاء الدين	'Alu'ēddin

When you are transliterating from the Latin script to the Arabic script and encounter a vowel with a macron (˘) above, you should immediately place a letter <ا> (Alef) in its place, if it was followed by two adjacent moon letters¹¹, you should also add a letter <ل> (Lam) after inserted <ا> (Alef).

6.1 Germination

In the Arabic script the germination can be denoted by shad-dah (ّ). In SLPO a germinated letter is written twice e.g. <پله> → <pelle>. Also notice that the second letter of two-letter Arabic loan words are always germinated regardless of their position in the sentence, e.g. <حدّ> → <hadd>, <حقّ> → <haqq> (Arabic words with duplicated root).

¹¹ Moon letters are: (ء), (ب), (ج), (ح), (خ), (ع), (غ), (ف), (ق), (ك), (م), (و), (ي).

(ه).

The geminated jim (جّ) in the Latin alphabet is represented by <dj> digraph. See table ?? for some examples.

Table 38: Examples for <dj>

Latin	Arabic
tavadjoh	توجه
motevadjeh	متوجه
sadjude	سجّاده
nadjur	نجّار
zadje	ضجّه

7 Orthography

After introduction of the letters, we are ready to discuss some issues related to orthography.

7.1 Punctuation

In this section, usages of hyphen and apostrophe in SLPO are described.

7.1.1 Hyphen

Hyphen usage is like its usage in English (in particular to join ordinarily separate words into single words to make some new technical or context-specific terms, but for general compound words no hyphen is used. See section ??).

A hyphen is also used in some interjection words, as you see in table ??.

Table 39: Interjection

Latin	Arabic	note
qol-qol	غُلْ غُلْ	<i>sound of boiling water</i>
teq-teq	تَغ تَغ	<i>clack</i>
qur-qur	قَار قَار	<i>croak</i>
tik-tuk	تِيك تَاك	<i>ticktack</i>
o-o	عَو عَو	<i>bark</i>
mio-mio	مِيُو مِيُو	<i>miaou</i>
cor-cor	شُر شُر	<i>gurgle</i>
baa-baa	بِع بِع	<i>bleat</i>

7.1.2 Apostrophe

Apostrophe has the following usages:

As a mark of contraction: Omission of one or more sounds (such as a vowel, a consonant, or a whole syllable) in a word or phrase to make easier pronunciation is very common in colloquial Persian and some times in Persian poetry. The omitted letters in a contraction are replaced by an apostrophe.

As a mark of elision: Elision has no influence on writing, but in poetry, theater text and other similar situations to show the actual speech of a character you can omit some characters and use an apostrophe instead.

Vocative form <'a> can be used to make vocative form nouns (vocative case). See table ???. Notice that the <'a> particle is pronounced /jɒ:/ if the word ends in a vowel and /ɒ:/ in other cases.

Table 40: Vocative form

Latin	Arabic
Xodu'u	خدایا
Parvardegur'u	پروودگارا
bozorg'u	بزرگا
Saadi'u	سعیدیا

7.2 Plural nouns

To form a plural noun add the <-hu> suffix to a singular noun. For example <sib>+<hu>→<sibhu> (<سیبها>).

Also for words that refer to humans you can form plural by adding the <-un> suffix. If the noun ends in a vowel it may be changed before adding <un>. See table ??? to see some examples [?].

Table 41: Using <an> to form plural form

کارمندان	karmand+un → karmandun
دانش آموزان	dunecamooz+un → dunecamoozun
معلمان	moallem+un → moallemun
استادان	ostad+un → ostadun
آشنایان	ucna+yun → ucnayun
بانوان	bunoo → bunov+ un → bunovun
خستوان	xastoo → xastov+un → xastovun

This suffix can also be added to some words to make new nouns that refer to a particular group. <giuh>+<un>→<giuhun>=<گیاهان> (refer to plants as a category).

For some Arabic loan words suffixes <-at> and <-in> can also be used to form plural nouns.

7.3 Ezafe

Ezafe has a wide range of uses in Persian, most importantly for making possessive and adjective phrases. [?, ?, ?], Ezafe should be written as a word in the Latin script: <e>. If the last letter of its previous word is a vowel, it is pronounced /je/, otherwise it is pronounced /e/. See table ??? and the following examples.

Table 42: Ezafe examples

کتابِ درسی	ketub e darsi
سیبِ شیرین	sib e cirin
خانه‌ی ویلایی	xune e vilui
آهویِ زیبا	uhoo e zibu
پرتویِ ماه	parto e muh
جویِ آب	joo e ub
جایِ خالی	ju e xuli

Example 1 *Mohammad ccuy ru dar goori e bozorg i ke az jens e rooy bood rixt va roo e miz gozuct.*

محمد چای را در قوری بزرگی که از جنس روی بود ریخت و روی میز گذاشت.

Example 2 *Be ju e negabduri e mohemmut zarrudxune mi goo'and.*

به جای نگهداری مهمات زرادخانه می‌گویند.

7.3.1 Comparison

You can use suffixes <-ar> (the "comparative") and <-tarin> (the "superlative") for comparison. Table ??? provides three examples.

7.4 Indefinite article

The Persian indefinite article is <i> that is placed after the referent. See table ??? for some examples. Note that in Arabic script the indefinite noun is formed by adding <ی> suffix and in Latin script it is a word.

Table 44: Indefinite article

definite		indefinite	
معرفه	maarefe	نکره	nakare
کتاب	ketub	کتابی	ketub i
کتاب‌ها	ketubhu	کتاب‌هایی	ketubhu i
خانه	xune	خانه‌ای	xune i
خانه‌ها	xunehu	خانه‌هایی	xunehu i

Example 3 *Har gerd i gerdoo nist.*
هر گردی گردو نیست.

Table 43: Comparison

Positive		Comparative		Superlative	
		تر-	-tar	ترین-	-tarin
بزرگ	bozorg	بزرگتر	bozorgtar	بزرگترین	bozorgtarin
زیبا	zibu	زیباتر	zibutar	زیباترین	zibatarin
پاک	puk	پاکتر	puktar	پاکترین	paktarin

Example 4 *Ali zarfi az zeytoon be man dud.*
علی ظرفی از زیتون به من داد.

Example 5 *Dunu i muru nejut dud.*
دانایی مرا نجات داد.
A wise person saved me.

Example 6 *Dunui maru nejut dud.*
دانایی مرا نجات داد.
The wisdom saved me.

The past stem of a lot of Persian verb can be built by adding <-id> to the end of the corresponding present stem. Additionally, the past participle can be formed by adding <-e> to the end of the past stem. To construct verbs, person suffixes are added to the three principal parts and if necessary some grammatical particles precede or secede the result. Verb suffixes (<می>, <نمی>, <ب> and <ن>) and person suffixes in the Arabic script, are written as a separate word in SLPO. Verbal person pronouns (person suffix in the Arabic script) are listed in table ??.

In table ?? you can see conjugation of indicative present of <raft an> (<رفتن>). Note that in the case of 3rd person singular, no suffix is added to the *past stem*; <ad> is added to *present stem* and <ast> is added to *past participle*.

7.5 Possessive Adjectives

Possessive suffixes in the Arabic Script are written as a word in SLPO. Please look at table ??

Table 45: Possessive Adjectives

ام	am	مان	mun
ات	at	تان	tun
اش	ac	شان	cun
Example 1			
کتابم	ketub am	کتابمان	ketub mun
کتابت	ketub at	کتابتان	ketub tun
کتابش	ketub ac	کتابشان	ketub cun
Example 2			
خانهام	xune am	خانهمان	xune mun
خانهات	xune at	خانهتان	xune tun
خانهاش	xune ac	خانهشان	xune cun

Table 46: Verbal Person Pronouns

Person	Singular	Plural
1st	am	im
2nd	i	id
3rd	φ ¹² / ad / ast	and

Table 47: Conjugation of indicative present <raft an> (<رفتن>)

Person	Singular	Plural
1st	mi rav'am	mi rav'im
2nd	mi rav'i	mi rav'id
3rd	mi rav'ad	mi rav'and

7.6 Verbs

All Persian Verbs in contemporary Persian can be built using three principal parts:

- Present stem
- Past stem
- Past participle

In table ?? examples of various tenses are presented.

¹²φ= nothing.

Table 48: Examples of various tenses of <did an>

خواهم دید	xvuh'am did
می‌خواهم ببینم	mi xvuh'am be bin'am
دارم می‌بینم	dur'am mi bin'am
دیده باشم	dide buc'am
دیده شدم	dide cod'am
دیده می‌شوم	dide mi cav'am

The general structure of Persian active verbs are demonstrated in figure ?? . Beginning from colored nodes (green nodes) and ending in dashed nodes, a verb group can be constructed. The figure can be used to form fifteen active types of Persian verbs. If a path contains any dashed edge ($--\rightarrow$), this means that type is not in active use any longer.

To construct the passive type counterpart of an active type, use <cod an> (<شدن>) in figure ?? and add the past participle part of the main verb to it. e.g. <duct'am mi did'am \rightarrow ducat'am dide mi cod'am>.

The general structure of Persian passive verbs are demonstrated in figure ?? . A verb can be constructed beginning from colored nodes (orange nodes) and ending in dashed nodes. The figure can be used to form 15 passive types of Persian verbs. If a path contains any dashed edge ($--\rightarrow$), this means that type is not in active use any longer.

A full description of Persian verbs is out of the scope of this paper. We will only show some examples here as a guide to write verbs in this orthography.

Table 49: Examples of three principal parts of some Persian verbs

Infinitive	Present stem	Past stem	Past Participant
بودن	buc	bood	boode
داشتن	dur	duct	ducte
خواستن	xvuh	xvust	xvuste
رفتن	rav	raft	rafte
شستن	cuy	cost	coste
زیستن	zi	zist	ziste
آمدن	u	umad	umade
اندیشیدن	andic	andicid	andicide
فهمیدن	fahm	fahmid	fahmide
دویدن	dav	david	davide
دواندن	davun	davunad	davunde
نامیدن	num	numid	numide
زنگیدن	zang	zangid	zangide

7.6.1 Infinitive form

The infinitive form of a verb is constructed by placing <an> after the past stem of the verb. e.g. <did>+< >+<an> \rightarrow <did an> (<دیدن>). e.g. "Lotfan feal e 'zist an' ru sarf kon."

Notice when it is used as a noun instead of infinitive, no space is placed between past stem and <an> suffix. e.g. "Ali az didan e tabiat xaste nemi cav'ad."

7.6.2 Imperative mood

Imperative is formed by preceding the present stem by <be> or <bi> particles (later used when the stem begins with a vowel).

Negated imperative is formed by preceding the present stem by <na> or in historical texts by <ma>. Table ?? shows some examples.

Table 50: Examples of imperative mood

	imperative		negated imperative
بخور	be xor	نخور	na xor
بنشین	be necin	ننشین	na necin
بیانید	bi andic	نیا نید	na andic
برو	be ro	نرو	na ro
ببین	be bin	نبین	na bin
بیا	bi u	نیا	na u
بشور	be coor	نشور	na coor
بخند	be xand	نخند	na xand
بفرما	be farmu	نفرما	na farmu
بده	be deh	نده	na deh

8 Ease of Acquiring

To measure how much time is required to learn SLPO, we conducted two experiments. In the first experiment we ask 14 native Persian speakers to read out a poem which was typeset in the Arabic script (each participant was isolated from others). After that we ask them to read the same poem written in SLPO. We counted the number of errors in their reading in both scripts. The figure shows the number of errors. It is obvious that the number of errors when reading a text in SLPO is dramatically less than reading the same text in the Arabic script. The test participants were trained less than five minutes to be familiar with SLPO. The age of the participants were between 12 and 50. Their education levels were between secondary school students and masters degree. Additionally, notice the errors in reading SLPO were the result of reading carelessly while the mistakes in reading the Arabic script were the result of its deficiencies.

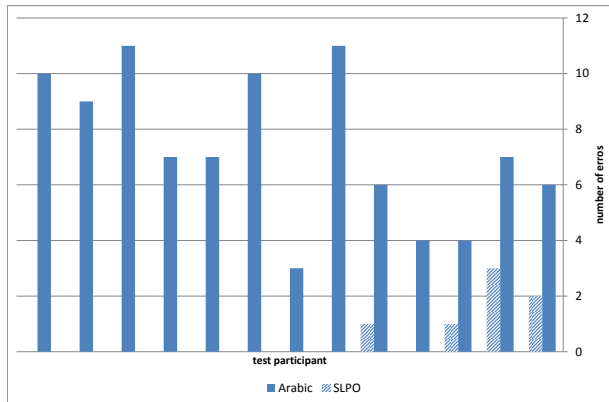


Figure 6: Total of errors among participants in the first experiment

To measure the time length required to learn SLPO >In the second experiment, we prepared a three-page manual for SLPO. We asked another 22 participants to read the manual and convert a given sentence in the Arabic script to SLPO and to convert another sentence in SLPO to its counterpart in the Arabic script. We requested from the participants to declare the time length it took them to read the manual and write down the answers. Analyzing the answers, we discovered they could map SLPO to the Arabic script perfectly. They also could convert the Arabic script to SLOP near-perfectly. In average, the time required for learning and using SLPO was 20.4 minutes among the participants.

9 Conclusions

In this paper we introduced a novel romanization and orthography scheme for Persian that is easy to learn and use, consistent and applicable in a lot of scenarios, particularly in writing academic materials (NLP and linguistics) and language learning. Here is a summary of its characteristics that make it superior to other proposed schemes:

(1) Using only the basic Latin alphabet that consequently makes SLPO usable in almost all computer environments without any configuration. (2) Persian has a lot of long words and phrases. Writing vowels in the Latin scripts make them even longer, to abate this issue: (i) Persian consonants and vowels are mapped to graphs and digraphs by considering their relative frequencies in the Persian language vocabulary (ii) using Persian language phonology rules and avoiding to write vowels in some situations, yet the pronunciation of words are completely identifiable by their spelling. As a result, the romanized text will be as short as possible (3) A clear and consistent rule is proposed to write the saken ayin. (4) The orthography scheme covers all Persian grammatical structures. (5) A complementary reversible romanization scheme is

provided for situations reversibility is required. Texts that are romanized by the reversible scheme, can be converted to the general scheme only by removing the diacritics. (6) According to our experiments, SLPO is easy to learn and use. A native speaker who has problems in reading classic Persian poetry can read them error-free when they are written in SLPO. Indeed SLPO can also help them to comprehend the poem more correctly. (7) The scheme has regular firm rules for writing verbs. (8) SLPO provides some clues to show some semantic aspects of texts. (9) A database of nearly all Persian words in SLPO is also available that can be easily used to create SLPO spell checker.

Acknowledgment

I would like to thank Prof. R. Safabakhsh, Dr. M. M. Homayounpour, Dr. S. Khadivi, Dr. M. Omidali, O. Ghayour, M. Kouhestani, A. Moloodi, M. Mahdavi, A. Mirzai to read the paper and to provide some suggestions, F. Ranjbaran to re-proof the paper, as well as all the other scholars who have helped in someways during devising SLPO. In addition, I would like thank all who participate in evaluation phase of SLPO.

References

- [1] Paarsi. <http://paarsi.org>.
- [2] Unipers. <http://www.unipers.com/>.
- [3] *Encyclopaedia Iranica Notes for Contributors*, 2012.
- [4] New persian romanization system. in *Tenth United Nations Conference on the Standardization of Geographical Names*, 2012.
- [5] Baluch, Bahman. Persian orthography and its relation to literacy. *Handbook of orthography and literacy*, pp. 365--376, 2006.
- [6] Cour, Jaafar. Sluh e mlu e fursi ru az koju corooa kon'im: be monusebat e tackil e farhangestun. *Yaqmu*, (273):157-161, 1971.
- [7] Farcidvard, Xosro. Bahs i darbure e xat e fursi va pishnahudhu i darbureye yeksunsun kardan e un. *Vahid*, (96):1318--1330, 1971.
- [8] Ftexuri, Seyyed Smuil. Darumad i bar xatcenusi va nazmgorizi e nevectur e fursi. *Dunecgade e oloom e nsuni dunecguh e emum Hoseyn*, (58):33--62, 2005.
- [9] Ghayour, Omid. Ironik. <http://xrad.ir/xat-e-ironik/>.

- [10] Hendi, Said. Dastoor e xat e fursi: Cive i dar negurec e kalamut e morakkab. *Rocd e zabun va adab e Fursi*, (63):27--31, 2002.
- [11] International Journal Of Middle East Studies. *IJMES Transliteration System For Arabic, Persian, And Turkish*.
- [12] Kahnemuyipour, Arsalan. Persian ezafe construction: case, agreement or something else. in *Proceedings of the 2nd Workshop on Persian Language and Computer, Tebran University, Tebran, Iran, 2006*.
- [13] Karimidoostun, Qolomhoseyn. Goonugooni e necunehu e jama '-ut' va '-un' dar fursi. *Oloom e jtemui va nsuni dunecguh e Ciruz*, (40):29--41.
- [14] Mahdavi, M. A. A proposed unicode-based extended romanization system for persian texts. *International Journal of Information Science and Management (IJISM)*, 10(1):57--71, 2012.
- [15] Maleki, Jalal. efarsi - a latin-based writing scheme for farsi.
- [16] Maleki, Jalal. A romanized transcription for persian. *Proceedings of Natural Language Processing Track (INFOS2008)*, Cairo, 2008.
- [17] Moslehabadi, Ali Moslehi. Ipaz (pársik). http://www.persiandirect.com/projects/ipa2/ipa2_tutor.htm.
- [18] Ostudi, Kuzem. Guhcomur e camsi e taqiir e xat e fursi (az qarn e avval e hejri tu konoon). *Payum e Baburestun*, 15, 2012.
- [19] Pantcheva, Marina Blagoeva. Persian preposition classes. 2006.
- [20] Rajabzude, Hucem. Andice e sluh va vasvase e taqir e xat dar irun va jupon e mouser. *Nume e Farbangestun*, (33):9--27, 2007.
- [21] Samiian, Vida. The ezafe construction: Some implications for the theory of x-bar syntax. *Persian Studies in North America*, pp. 17--41, 1994.
- [22] Vahedi-Langrudi, Mohammad Mehdi. Lexical representation of morphologically complex words and lexical access in persian. *Zabun va adabiut e Fursi*, (33):35--66, 2001.

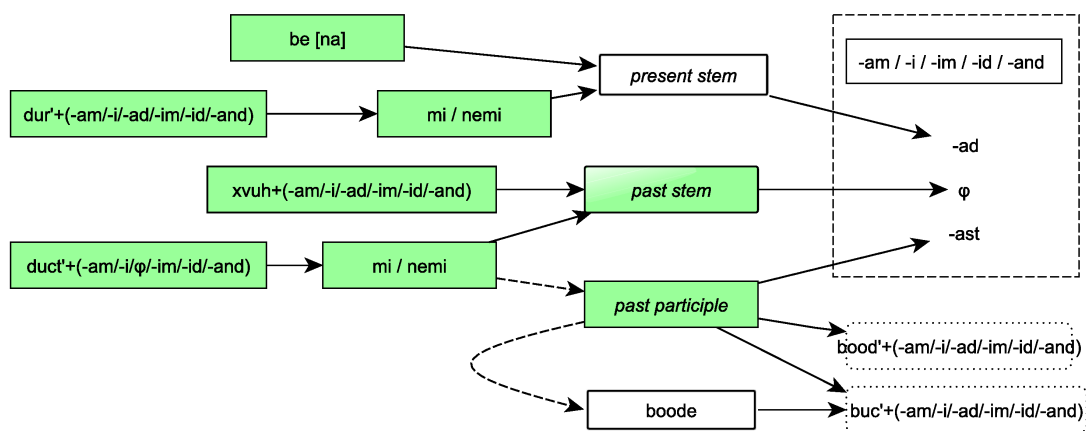


Figure 4: General pattern of Persian active verbs

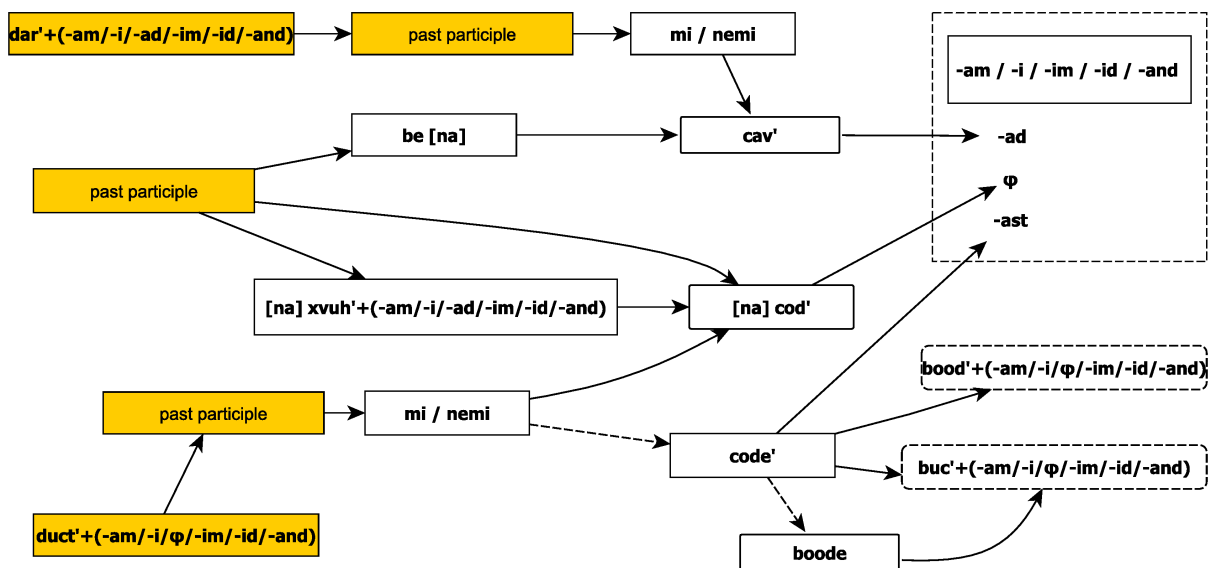


Figure 5: General pattern of Persian passive verbs

A Additional Examples

In this section we provide some important groups of words to be used as a reference and a learning material.

A.1 Name of months

Month's names are capitalized in the orthography. Table ?? shows the names of Solar Hijri months in Latin script. Month's names and their 3-letter abbreviation are shown in table ??

Arabic	Latin	Latin Abr.
فروردین	Farvardin	Frv
اردیبهشت	Ordibehect	Ord
خرداد	Xordud	Xrd
تیر	Tir	Tir
مرداد	Mordud	Mrd
شهریور	Cahrivar	Chr
مهر	Mehr	Mhr
آبان	Ubun	Ubn
آذر	Uzar	Uzr
دی	Dey	Dey
بهمن	Bahman	Bhm
اسفند	Sfand	Sfn

A.2 Week days

Names of weekdays should be capitalized in SLPO. Table ?? shows the names and their 1-letter and 3-letter abbreviations.

Arabic	Latin	1-letter abr.	3-letter abr.
شنبه	Canbe	C	Cnb
یکشنبه	Yekcanbe	Y	Ykc
دوشنبه	Docanbe	D	Doc
سه‌شنبه	Secanbe	S	Sec
چهارشنبه	Ccurcanbe	R	Ccu
پنج‌شنبه	Panjanbe	P	Pnc
آدینه/جمعه	Udine/Jome	U / J	Udn / Jom

A.3 Pronouns

A.3.1 Subject pronouns

Table 53: Subject pronouns

من	man	ما	mu
تو	to	شما	comu
او	oo	ایشان	icun

A.3.2 Object Pronouns

Table 54: Object pronouns

مرا	maru	مارا	mura
ترا	toru	شمارا	comuru
اورا	ooru	ایشانرا	icunru

A.4 Demonstrative pronouns

Table 55: Demonstrative pronouns

این	in
آن	un
این‌ها	inhu
آن‌ها	unhu
اینان	inun
آنان	unun

<Inan> and <anan> are usually used only for humans.

A.5 Interrogative words

Table 56: Interrogative words

چه (چی)	cce (cci)	what
چرا	cceru	why
چگونه	ccegoone	how
چند	ccand	(how much)
چندتا	ccandtu	how many
چقدر	cceqadr	how much
کی	ki	who
کی	key	when
کجا	koju	where

A.6 Numbers

A.6.1 Cardinal numbers

Table ?? lists cardinal numbers. As an example 684,541,213,251 in letters is:

*cecsad o hactud o ccur bilyun o
punsad o ccel o yek milioon o
devist o sizdah hezur o
devist o panjuh o yek.*

A.6.2 Ordinal numbers

Table ?? contains the first five ordinal numbers.

Table 58: Ordinal numbers

English	number+om		number+min	
1st	1om	yekom/naxost	1mn	naxostin
2nd	2om	dovom	2mn	dovomin
3rd	3om	sevom	3mn	sevomin
4th	4om	ccurom	4mn	ccuromin
5th	5om	panjom	5mn	panjmin

A.7 Poetry

Tables ?? and ?? represent two poems by Hufez and Saadi.

Table 57: Cardinal numbers

0	sefr
1	yek
2	do
3	se
4	ccur ¹³
5	panj
6	cec
7	haft
8	hact
9	noh
10	dah
11	yuzdah
12	davuzdah
13	sizdah
14	ccurdah
15	punzdah
16	cunzdah
17	hefdah
18	hejdah
19	noozdah
20	bist
21	bist o yek
:	:
29	bist o noh
30	si
40	ccehel
50	panjuh
60	cast
70	haftud
80	hactud
90	navad
100	sad
200	devist
300	sisad
400	ccursad
500	punsad
600	cecsad
700	haftsad
800	hactsad
900	nohsad
1,000	hezur
1,000,000	milioon
1,000,000,000	bilyoon/bilyurd
1,000,000,000,000	tirilyoon

¹³Also <ccehur>

Table 59: An example of poetry in Latin script (by Saadi)

بنی آدم اعضای یکدیگرند	Baniudam aazu e yekdigar'and
که در آفرینش ز یک گوهرند	Ke dar ufarinec ze yek gohar'and
چو عضوی به درد آورد روزگار	Cco ozv i be dard uvar'ad roozgar
دگر عضوها را نماند قرار	Degar ozvhu ru na mun'ad qarur
تو کز محنت دیگران بی غمی	To k'az mehnat e digarun biqam'i
نشاید که نامت نهند آدمی	Na cuy'ad ke num at nah'and udami

Table 60: An example of poetry in the general and reversible schemes (Hafez's 64th sonnet)

زبان خموش ولیکن دهان پر از عربیست	اگرچه عرض هنر پیش یار بی ادبیست
Agarce arz e honar pic e yur biaadabi'st	Zabun xamooc valikan dahun por az Arabi'st
Agarce 'arz e honar pic e yur biaadabi'st	Zabun xamooc valikan dahun por az 'Arabi'st
بسوخت دیده ز حیرت که این چه بوالعجبیست	پری نهفته رخ و دیو در کرشمه‌ی حسن
Pari nahofte rox o div dar kerercme e hosn	Be sooxt dide ze heyrat ke in cce bolajab i'st
Pari nahofte rox o div dar kerercme e ḥosn	Be sooxt dide ze ḥeyrat ke in cce bol'ajab i'st
چراغ مصطفوی با شرار بولهیبیست	در این چمن گل بی خار کس نجید؛ آری
Dar in ccaman gol e bixur kas na ccid; uri	Cceruq e mostavafi bu carur e boolahabi'st.
Dar in ccaman gol e bixur kas na ccid; uri	Cceruq e moštavafi bu carur e boolahabi'st.
که کام بخشی او را بهانه بی سببیست	سبب مپرس که چرخ از چه سفته پرور شد
Sabab ma pors ke ccarx az cce sefleparvar cod	Ke kumbaxci e oo ru bahun e bisababi'st
Sabab ma pors ke ccarx az cce sefleparvar cod	Ke kumbaxci e oo ru bahun e bisababi'st
مرا که مصطبه ایوان و پای خم طنیبیست	به نیم جو نخرم طاق خانقاه و رباط
Be nim jo na xaram tuq e xunequh o rebut	Maru ke mastabe eyvun o pu e xom tanab'ist.
Be nim jo na xaram tuq e xunequh o rebuṭ	Maru ke maštabe eyvun o pu e xom ṭanab'ist.
که در نقاب زجاجی و پرده‌ی عنیبیست	جمال دختر رز نور چشم ماست مگر
Jamul e doxtar e raz noor e ccacm e mu'st magar	Ke dar nequb e zojuji o parde e enabi'st.
Jamul e doxtar e raz noor e ccacm e mu'st magar	Ke dar nequb e zojuji o parde e 'enabi'st.
کنون که مست خرابم صلاح بی ادبیست	هزار عقل و ادب داشتم من؛ ای خواجه
Hezur aql o adab ductam man ey xvuje	Koonon ke mast e xarub'am saluh biadabi'st.
Hezur 'aql o adab ductam man ey xvuje	Koonon ke mast e xarub'am ṣaluh biadabi'st.
به گریه‌ی سحری و نیاز نیم شبیست	بیار می که چو حافظ هزارم استظهار
Biur mey ke cco Hufez hazur'am stezhur	Be gerye e sahari o niuz e nimcab i'st.
Biur mey ke cco Ḥufez hazur'am stezhur	Be gerye e sahari o niuz e nimcab i'st.