

MP080034

MITRE TECHNICAL REPORT

# The Structure of Persian Names

**February 2008**

Karine Megerdooian

**Contract No.:** W15P7T-07-C-F  
**Project No.:** 0707N7AZ-0C/1C

The views, opinions and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

©2008 The MITRE Corporation. All Rights Reserved.

**MITRE**

## **Abstract**

This report provides a description of the structure of common Persian names from Iran, with an emphasis on automatic recognition of these personal names. Since the early parts of the 20<sup>th</sup> century, person proper names of Persian (Farsi) origin have been composed of a first name and a last name. There are no middle names. However, first and last names may each be a compound proper name, consisting of two subparts. The report discusses the various components of person names such as titles, honorifics, the internal structure of the first and last name, as well as affixes used to form the latter. The characteristics of the Perso-Arabic writing system (lack of capitalization, absence of short vowels, optionality in spacing) and variability in the English transcriptions are discussed in more detail as they may give rise to a number of issues for NLP applications of name matching.



# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>First Names</b>	<b>1</b>
2.1	Simple Forms	2
2.1.1	Islamic names	2
2.1.2	Nouns or adjectives of Persian origin	2
2.1.3	Persian literary and historical figures	2
2.2	Compound Forms	2
2.2.1	Two Islamic or Arabic names	2
2.2.2	“Qolam” or “Gholam” names	2
<b>3</b>	<b>Last Names</b>	<b>3</b>
3.1	Simple Forms	3
3.2	Affixal Forms	3
3.2.1	Suffixes	3
3.3	Prefixes	5
3.4	Names ending in –ian or –yan:	5
3.5	Compound Forms	6
<b>4</b>	<b>Titles and Honorifics</b>	<b>6</b>
<b>5</b>	<b>Orthographic and Phonological Notes</b>	<b>7</b>
5.1	Variability in English transcription	7
5.2	Lack of capitalization	8
5.3	Spacing issues	8
<b>6</b>	<b>References</b>	<b>10</b>

## List of Tables

**Table 1:** Orthography variation in Persian last names

8

# 1 Introduction

Prior to the 20<sup>th</sup> century, person proper names in Iran did not include a surname and people were often distinguished by their place of birth, profession and honorific titles. In the early 1920s, the secularization and modernization policies of the government of Reza Shah Pahlavi required the use of surnames. Family names were selected relating to geographic regions, professions, or by using abstract concepts that depict a positive human trait.

Modern proper names of Persian origin are composed of a first name and a last name. There are no middle names. However, each component can be a compound proper name, consisting of two subparts. This report gives a description of the structure of common Persian names from Iran.

**Section 1: First names.** First names may be of Arabic origin, usually related to Islamic themes. These names may follow the internal structure of Arabic names such as *Abdolrashid* (=Abd+Al+Rashid). Mostly, however, they are simple or compound forms such as *Mohammad* or *Mohammad Reza*, respectively. There are also many first names of Persian origin, such as *Kiavash* or *Parastoo*.

**Section 2: Last names.** The most common ending for Persian last names is the “-i” suffix. Last names can also appear with a number of affixes of Persian origin, or without any affixes at all. Last names may also be in compound form, resulting from the juxtaposition of two simple last names.

**Section 3: Titles and honorifics.** Certain titles, specially religious ones, behave as part of the proper name. These include terms like *Haji* or *Seyyed*.

**Section 4: Orthography.** The writing system (lack of capitalization, absence of short vowels, optionality in spacing) and variability in the English transcriptions give rise to a number of ambiguities that may raise problems for NLP applications of name matching.

This report focuses on proper names originating in Iran and does not cover the structure of Persian names in Afghanistan and Tajikistan. In addition, the proper name structures of the various ethnic groups living in Iran (e.g., Kurds, Azeris, Baluchis) are not described in this report.

## 2 First Names

First names can be either simple (e.g., Anousha, Maryam, Behzad, Ahmad) or compound (e.g., Mohammad Mehdi, Ali Reza, Amir Hossein).

## 2.1 Simple Forms

### 2.1.1 Islamic names

The most common first names are names of Islamic origin, especially Shiite ones.

*Examples: Mohamad, Reza, Ali, Hossein, Hassan, Mehdi, Fatemeh, Zahra, Said*

Some of these names follow the Arabic naming patterns.

*Examples: Abdollah (Abd+Allah), Ruhollah (Ruh+Allah), Abolghassem (Abu+al+Ghassem), Abolfazl (Abu+al+Fazl), Ashaollah (Asha+Allah), Abdolrashid (Abd+al+Rashid), Najmeddin (Najm+al+Din)*

### 2.1.2 Nouns or adjectives of Persian origin

Most Persian names have a meaning. These names are generally nouns or adjectives being used as proper names.

*Examples: Aram (calm), Nader (rare), Mahtab (moonlight), Yasaman (jasmine), Parastu (swallow), Shirin (sweet)*

However, the original word may have fallen out of usage in the language, with the proper name usage becoming the main one.

*Examples: Afshar (companion), Rastin (true, truthful), Kamnoosh (lit. sweet palate; happy fruition)*

### 2.1.3 Persian literary and historical figures

These include the names of legendary figures, kings and religious personalities, as well as old gods and goddesses.

*Examples: Cyrus, Dariush, Kaveh, Mitra, Mani, Fereydoon*

## 2.2 Compound Forms

Persian first names can appear as compounds, which are formed by combining two first names (usually both of Arabic or Islamic origin):

### 2.2.1 Two Islamic or Arabic names

*Examples: Mohamad Naser, Mohammad Ali, Mohamad Taghi, Amir Mansur, Mohamad Reza, Alireza, Mohamad Hossein, Mostafa Mohamad, Hamidreza, Mohamad Bagher*

### 2.2.2 “Qolam” or “Gholam” names

Certain compound names are formed with Gholam meaning “servant”.

*Examples: Qolamhossein, Qolamreza, Gholamreza, Qolam Heydar*

## 3 Last Names

Last names can also appear as simple forms (e.g., *Parsi*, *Ganji*) or compound forms (e.g., *Karimi-Hakkak*, *Vahedi Langarudi*, *Shariat Panahi*). In addition, Persian last names may contain prefixes and suffixes that may be written attached or as a separate word (e.g., *Mirkazemi*, *Ahmadinejad*, *Aryanpur*, *Amin Far*).

### 3.1 Simple Forms

Last names may appear without any affixes making them indistinguishable from regular nouns and adjectives<sup>1</sup>:

**Examples:** *Mofid* (useful), *Azad* (free), *Khayat* (tailor), *Deldar* (sweetheart), *Bozorgmehr* (of great affection), *Hedayat* (guidance), *Pezeshk* (doctor), *Zare* (farmer)

### 3.2 Affixal Forms

The most common affixes used in forming Persian last names are listed below. *Note:* Almost all of these affixes can be written, in the original Persian script, as attached to the main part or they can appear in detached form. In the English transcription, these affixes may also be written as a detached word. This will be discussed in more detail in Section 45.

#### 3.2.1 Suffixes

- **-i, -y, -ee**

This is the most common suffix used in the formation of last names and it can be:

(a) Attributive, formed on a noun or adjective. Note that the root can be of either Persian or Arabic origin (including Arabic compounds such as ‘*Ruholamin*’)

**Examples:** *Karimi* (generous+i), *Ganji* (treasure+i), *Mazloomi* (meek+i), *Darvishi* (dervish+i), *Shojaee* (brave+ee), *Golesorkhi* (rose+i), *Zahedi* (devout+i), *Nafisi* (precious+i), *Ruholamini* (honest soul+i), *Sharifi* (noble+i)

(b) Geographical, referring to a region or place (usually of birth or residence of the family ancestors)

**Examples:** *Tehrani*, *Shirazi*, *Hamadani*, *Khomeini*, *Sari*, *Zanjani*, *Shemirani*, *Rafsanjani*, *Boushehri*, *Yazdi*, *Esfahani*, *Dezfouli*, *Araghi*, *Asfarjani*, *Shahsavari*, *Kouhestani*

(c) Occupational, referring to a trade or profession

**Examples:** *Tabibi* (doctor+i), *Kalantari* (sheriff+i)

(d) Other attribution, usually based on a first name that often refers to a legendary or

---

<sup>1</sup> Note that there is no capitalization in Persian script.

historical figure.

*Examples: Jamshidi, Yusefi, Siavashi, Sohrabi*

- **-abadi** (= village)  
*Ahmadabadi, Dowlatabadi, Soltan Abadi, Azeem Abadi, Khorramabadi*
- **-ara** (= adorer)  
*Golara, Razmara, Saleh Ara, FarmanAra, Chamanara, Delara*
- **-chi:** of Turkish origin, indicating geographic location or profession  
*Maragheh chi, Chayi chi, Golabchi, Ahanchi, Gheitanchi*
- **-dust** (= one who likes)  
*Rahmandust, Jahandust, Haghghat Dust*
- **-fard** (= individual, unique)  
*Azmayesh Fard, Heydaryan Fard, Danayifard, Ahmadifard*
- **-far** (= splendor)  
*Kiyanfar, Niafar, Omidifar, Hesamfar, Rezvanfar, Tabatabayifar*
- **-ian, -yan** (= of the family of)  
[See note below]
- **-ju** (= searcher of)  
*Ashtiju, Hagh Jou, Haghghatjoo*
- **-kiya, -kia** (= defender, king)  
*Saidi Kia, Arefkia, Hatami Kia, Mahdavikiya*
- **-manesh** (= of disposition of)  
*Azadmanesh, Radmanesh, Sohrab Manesh, Fardmanesh, Nikmanesh, Arya Manesh*
- **-mand** (=possessing)  
*Arjmand, Alaghmand, Firoozmand*
- **-niya, -nia** (= ancestor)  
*Tajernia, Pirniya, Behroozinia, Rahim Nia, Sharifi Nia, Safinia, Ashoorinia*
- **-nejad, -nezhad, -nejhad** (= race)  
*Ahmadinejad, Iran Nejad, Abbasinejad, Niknezhad, Verdinejhad, Saminejad, Nabavi Nejad*
- **-parast** (= admirer of)  
*Ayinparast, Akbari Parast, Haghaniparast, Din Parast*
- **-pur, -pour, -poor** (= son of)  
*Mirzapur, Kahnemuyipour, Radpour, Behzadpour, Rastinpur, Gharibpour, Aryanpur, Farahani Pour, Zandipour, Ferdowsipour, Iranpour, Darvishpoor*
- **-rad** (= brave, gentleman)  
*Ensarirad, Arvin Rad, Mansouri Rad, Mousavirad*
- **-tabar** (= of family of)  
*Jafari Tabar, Hajitabar, Alavitabar, Rafiei Tabar, Kazemitabar*

- **-vand**

*Saberivand, Rostami Vand, Esmailvand, Aminvand, Morteza vand*

- **-var** (= of profession of, agent of)

*Pisheh var, Daneshvar*

- **-yar** (= friend)

*Mehryar, Ahmadiyar, Nikyar*

- **-zadeh** (= born of)

*Bagherzadeh, Alizadeh, Soltanzadeh, Bahram Zadeh, Hasanzadeh, Mehranzadeh, Mohammadzadeh, Koochakzadeh, Sharifzadeh, Akhoundzadeh, Mojtahedzadeh*

- **-zad** (= born of)

*Javad Zad, Farrokhzad, Hoseinizad, Khalil zad*

### 3.3 Prefixes

- **kia-** (= defender, king)

*Kiarostami*

- **mir-** (= master of)

*Mirdamadi, Mir Mohammadi, Mirkazemi, Mir Masoumi, Mirlatifi, Mirabbaasi*

- **pur-, pour-** (= son of)

*Purshaghghi, Pour Bonab, Pourmohsen, Purhadi*

- **yar-** (= friend)

*Yarmohammadi, Yarshater*

### 3.4 Names ending in –ian or –yan:

Certain names of Persian origin end in “ian” or “yan”. However, most last names that end in this suffix are traditionally Armenian last names. The distinction has to do with the root form of the last name: If the root element is of Persian origin (and not of Armenian origin), then the last name is most probably a Persian name.

The following are examples of last names ending in “ian” which are of Persian origin:

<i>Tabatabayan</i>	→ <i>tabatabayi</i> = descendent from prophet on both sides
<i>Nojumyan</i>	→ <i>nojum</i> = astronomy
<i>Beheshtian</i>	→ <i>behesht</i> = paradise
<i>Kazemian</i>	→ <i>kazem</i> = a first name; one who represses his anger
<i>Farmanfarmayan</i>	→ <i>farmanfarma</i> = governor-general
<i>Goldoozian</i>	→ <i>goldooz</i> = one who sews/embroiders flowers

These are in contrast with the following Armenian last names, which are based on words or names of Armenian origin:

*Stepanyan, Megerdichian, Kaputikian, Harutunyan, Grikorian, Melik Hagopian, Ter Nersessian, Der Kureghian*

Some last names ending in “ian” (or “yan”) may be ambiguous, however. These are names that have roots of Persian origin yet have historically been used by Armenians as well:

<i>Khodabakhshian</i>	→	<i>khoda</i> = god + <i>bakhsh</i> = offering
<i>Dabbaghian</i>	→	<i>dabbagh</i> = tanner
<i>Gharibyan</i>	→	<i>gharib</i> = stranger, foreigner
<i>Eskandaian</i>	→	<i>eskandar</i> = a first name (Alexander)
<i>Teymoorian</i>	→	<i>teymoor</i> = a first name (Timur)

### 3.5 Compound Forms

The last names may consist of two components as shown, where either the first or second component ends in the –i suffix:

*Examples: Danesh Jafari, Vaez Javadi, Rahim Moshayi, Darya Bandari, Shariat Panahi, Moshfegh Kashani, Tabibzadeh Nouri, Majnoon Hoseini, Sabokdast Noudehi, Khayyat Khalighi*

*Examples: Karimi Hakkak, Kharrazi Pakdel, Moradi Shahrabak, Hayati Moghaddam*

Both parts may end in the –i affix as in the following examples. In these instances, the second component generally refers to a geographic location.

*Examples: Bagheri Lankarani, Vahedi Langarudi, Akhani Sanjani, Habibi Rezaiee, Yazdi Samadi, Esmaili Sardari*

## 4 Titles and Honorifics

The following honorifics are used either before or after a proper name:

- **Agha** آقا - Sir, mister. General term of respect.
- **Ayatollah** آیت الله - “Sign of God”, high title given to major Shia clergymen.
- **Darvish (Dervish)** درویش - A Sufi mystic.
- **Haji** حاجی - Pilgrim; a man who has made the Muslim pilgrimage to Mecca.
- **Haj Agha** حاج آقا - Form of address for a male pilgrim.
- **Haj Khanom** حاج خانم - Form of address for a female pilgrim.
- **Hojatoleslam** حجت الاسلام - Eminence; an honorific title meaning "proof of Islam", given to middle-ranking Shia clerics of the rank of mujtahid.
- **Imam (Emam)** امام
- **Karbala'i** کربلائی - Before the name, one who has made the pilgrimage to Karbala. Following the name, one from Karbala.
- **Khan** خان - Following the name, served at one time as a title, but now considered merely as an honorific.
- **Mashhadi** مشهدی - Before the name, one who has made the pilgrimage to Mashhad. Following the name, one from Mashhad.

- **Mirza** میرزا - Before the name, a general term of respect which usually indicates that the one designated is literate. Following the name, it indicates a man of royal descent.
- **Mulla** ملا - A Muslim priest.
- **Ostad** استاد - A master craftsman.
- **Seyyed** سید - A descendant of the Prophet Muhammad.
- **Seyyede** سیده - A female descendant of the Prophet Muhammad.
- **Shaykh** شیخ - An elder; a chief; a professor; or the head of a dervish order.

*Note:* Some of these titles such as *Darvish* or *Mirza* are not used as often in modern times, however the Islamic terms such as *Haji* or *Seyyed* are more prominent these days (i.e., since the establishment of the Islamic Republic). *Haj Agha* and *Haj Khanoom* are used as general forms of address these days (even if the person addressed has not been to Haj).

The list of titles and honorifics mentioned here are far from exhaustive and they may be recursive, as the following example illustrates:

*Hojatoleslam ol-moslemin Haj Seyyed Ahmad Khomeyni*  
 حجت‌الاسلام والمسلمین حاج سید احمد خمینی

## 5 Orthographic and Phonological Notes

Persian uses an extended version of the Arabic script. In this writing system, short vowels are often not represented allowing for a number of ambiguities in English transcription. There is no capitalization in this script to distinguish proper names from basic nouns or adjectives. There is further ambiguity due to the fact that there are many ways of transcribing names from the Persian script into English. In addition, affixes on last names may be written either in attached form or as a separate word in the original script, giving rise to more variability in transcription. These issues are discussed in more detail below.

### 5.1 Variability in English transcription

Arabic and Persian names can be transcribed in a number of ways into English (Latin) script. For instance, the proper name محمد can be written as *Mohammad*, *Mohamad*, *Mohammed*, *Muhammad*, etc. Similarly, the name نجم الدين can be transcribed as *Najmeldin* or *Najm-al-din* based on the original orthography, or it can be written as *Najmeddin* to represent the pronunciation of the name.

In general, the issues that arise in transcription of Arabic names into English also apply to the transcription of Persian proper names. It should be noted, however, that the pronunciation of names of Arabic origin is different in Persian and therefore the transcription should also be modified to represent the Persian names. For instance, the sound /w/ does not exist in the Persian of Iran and instead /v/ is used; hence, what would be transcribed as *Dawood* or

*Dawud* or *Da'oud* based on the Arabic pronunciation should now be represented as *Davood* or *Davud*.

## 5.2 Lack of capitalization

As mentioned in the first two sections, most proper names in Persian are actually nouns or adjectives. Since there is no capitalization in the writing system, this gives rise to ambiguities in detecting proper names from text.

## 5.3 Spacing issues

In the Persian writing system, a space is usually inserted between the first and last names. However, affixes may appear without an intervening space. In these instances, if the name root ends in a character that has a final form<sup>2</sup>, either the affix and the name root are attached to each other or they remain in detached form but with an intervening control character, namely the *zero-width non-joiner* (unicode: \u200c). This is illustrated in more detail in the examples below for last names ending with the suffix “far”.

	Description	Persian script	Transcription
<b>Name root ending in character without final form</b>	No intervening space between suffix ‘far’ and root part <i>Pirooz</i> . The character ‘z’ only has one form (i.e., no attached/detached distinction).	پیروزفر	Piroozfar
<b>Name root ending in character with final form</b>	No intervening space between suffix ‘far’ and root part <i>Kamal</i> . The character ‘l’ of <i>Kamal</i> appears in attached form.	کمالفر	Kamalfar
	Space between suffix ‘far’ and root part <i>Ramin</i> . The character ‘n’ of <i>Ramin</i> appears in detached or final form.	رامین فر	Raminfar
	No space but an intervening control character between suffix ‘far’ and root part <i>Ramin</i> . The character ‘n’ of <i>Ramin</i> appears in detached or final form.	رامین‌فر	Raminfar

**Table 1:** Orthography variation in Persian last names

This variability in the attachment of affixes is also represented in the English transcription of Persian names.

<sup>2</sup> Characters that do not have a final form are *alef*, *dal*, *zal*, *re*, *ze*, *zhe*. These characters do not change form. All other letters may appear in attached form or in final form, depending on the context (same as in Arabic).

<b>Examples:</b>	<i>Abbasi Nejad</i>	vs.	<i>Abbasinejad</i>
	<i>Chayi chi</i>	vs.	<i>Golabchi</i>
	<i>Mir Damadi</i>	vs.	<i>Mirdamadi</i>
	<i>Pour Bonab</i>	vs.	<i>Pourmohammadi</i>

Similarly, compound last names may be transcribed as a single word or as a compound word in English (e.g., *Nik Khah* vs. *Nikkhah*)

In addition, certain transcriptions of Persian names allow hyphens inbetween the affix and the name root, or between the two subparts of compound names:

<u>Affix and Root:</u>	<i>Naghshineh-Pour, Pour-Mohammadi, Hosseini-Nejad</i>
<u>Compound first names:</u>	<i>Mohammad-Reza, Gholam-Hossein</i>
<u>Compound last names:</u>	<i>Najafi-Shahri, Haji-Abdolhosseini, Malek-Mohammadi, Savaghebi-Firoozabadi, Habibollah-Eskandari</i>

Finally, the compound forms in Persian are linked in pronunciation by the ‘ezafe’, an affix that generally relates the components of a noun phrase to each other. This affix is pronounced as /e/ and is usually an unwritten vowel. Sometimes, the transcription of the Persian name may represent the pronunciation as it is in Persian, as in *Majnoon-e-Hosseini*. This is, however, quite rare and in most transcriptions the ‘ezafe’ vowel is not represented in the English transcription.

**Some more examples:** These examples include compound first names and last names in various forms.<sup>3</sup>

- Mahmoud Ahmadinejad
- Hossein Rahim Pour Bonab
- Mohammad Javan Nik Khah
- Ali Akbar Shah Nejat Booshehri
- Shir mohammad Moez Ardalan
- Gholam Reza Rasoulian
- Ali Nikkhah
- Seyyed Reza Mirayi Ashtiani
- Mohammad-Reza Najafi-Shahri
- Gholam-Reza Savaghebi-Firoozabadi
- Seyyed Gholam Reza Eslami
- Seyyed Bagher Mirabbasi
- Iraj Malek-Mohammadi

---

<sup>3</sup> Except for the first name in this list, the rest are professors at the University of Tehran. The names were found under the department faculty pages at <http://public.ut.ac.ir/html/activ.html>.

## 6 References

The Persian-American Yellow Pages. Washington DC, Virginia, Maryland areas.

[http://en.wikipedia.org/wiki/Persian\\_names](http://en.wikipedia.org/wiki/Persian_names)

<http://www.princeton.edu/~nasab/typesur.html>

<http://www.answers.com/topic/family-name>

[http://bahai-library.com/index.php5?file=wilmette\\_persian\\_arabic\\_names](http://bahai-library.com/index.php5?file=wilmette_persian_arabic_names)

Searches on [www.google.com](http://www.google.com)

[www.farsidic.com](http://www.farsidic.com)