

D R A F T - D R A F T - D R A F T - D R A F T - D R A F T
D R A F T - D R A F T - D R A F T - D R A F T - D R A F T

eFarsi - A Latin-Based Writing Scheme for Persian

Jalal Maleki

Department of Computer and Information Science
Linköping University
SE-581 83 Linköping
Sweden

Email: jma@ida.liu.se

December 2003

D R A F T - D R A F T - D R A F T - D R A F T - D R A F T
D R A F T - D R A F T - D R A F T - D R A F T - D R A F T

Abstract

This paper proposes a writing scheme for transliteration of Persian to Latin. The scheme, called eFarsi, introduces an alphabet and a number of transliteration conventions that provide a practical and easy to learn writing method.

Keywords: *Farsi, Persian, Orthography, Romanization, Transliteration, Arabic Script, Persian-Arabic Script*

Contents

1	Introduction	2
1.1	Persian-Arabic Script	4
1.2	General Characteristics of eFarsi	5
2	eFarsi Alphabet	6
3	eFarsi Conventions	16
3.1	Writing Words As Pronounced	16
3.1.1	Some Arabic-Related Issues	17
3.2	<i>Ist</i>	19
3.3	No Silent <i>h</i>	20

3.4	<i>Taâdid</i>	20
3.4.1	<i>Taâdid</i> on the Persian-Arabic letter <i>Ye</i>	21
3.4.2	<i>Taâdid</i> on the final letter	21
3.5	Ezâfe	21
3.6	Compound Words	24
3.6.1	Constituents of Compound Words	24
3.6.2	Writing Compound Words	26
3.6.3	Exceptions	31
3.7	Geographic Names, Proper Names and Trademarks	34
3.8	Foreign words	34
3.9	Some Modifications and Extensions	36
3.10	Making eFarsi More Practical	36
3.11	Improving the look of the words	36
3.11.1	<i>Taâdid</i> on Certain Diphthongs	37
3.11.2	<i>Taâdid</i> on <i>Ist</i>	37
3.11.3	<i>Ist</i> Between Vowels	37
3.12	Capitalization, Abbreviation and Punctuation	38
3.12.1	Capitalization	38
3.12.2	Abbreviation	39
3.12.3	Punctuation	41
4	Common Mistakes	42
5	Discussion and Further Work	43
5.1	Why Latin?	43
5.2	Further Issues	44

1 Introduction

This paper introduces an orthographic scheme for transliteration of Farsi to Latin. This scheme, called eFarsi, consists of an alphabet and a number of conventions that provide a practical and easy to learn writing method. Currently, Persian is written in a variation of the Arabic script that we will refer to as the *Persian-Arabic Script* (PA-Script hereon) which is mainly used in Iran. Similar scripts are used in Afghanistan, Pakistan and parts of India. Persian is currently written in Cyrillic script as well (Tajikestan). A Latin script was introduced in Tajikestan in the 1920's but was soon abandoned in favor of the Cyrillic script. Before going further we will write the first two sentences of this introduction in eFarsi.

Dis peyper interodiuses an ortogerafik eskim for teranslitereyşen âv Fârsi tu Latin. Dis eskim, kâld eFarsi, kânsists âv an alfâbet and e nâmbur âv kânvensens dat perovâyde e peraktikâl and izi tu lern râyting metod.

We are introducing eFarsi for various reasons. The main reason being that there are many people worldwide that are able to speak Persian but do not have good command of the PA-Script. There are also millions of Farsi speakers who do not have access to Persian-Arabic fonts on their computer keyboards and even if they did, are usually not able to employ them correctly and effectively in the variety of frequently upgraded software tools commonly used for communication and text and graphic processing. It is also our understanding that a large number of Persian speakers use some sort of Latin-based transliteration for sending SMS, email and chatting on the Internet but each individual writes in a different way. Introduction of a Latin-based alphabet for Farsi is not a new idea. Since the beginning of the 20th century there have been a number of proposals for the romanization of Persian script [5]. In [7], Lambton briefly describes a transliteration scheme for Persian and uses it throughout his grammar book. There are also a number of recent proposals such as Eurofarsi [3] and Unipers [4] which are mainly published on the WWW. Unfortunately, there is no standard scheme and most of the official attempts initiated by the United Nations have been limited to standardization of the geographic names in the Persian-speaking countries.

As well as hoping to contribute towards a standard for writing Persian in Latin which will drastically improve the quality of the electronic information, we hope that eFarsi would facilitate communication between Persian speakers and also encourage others to learn Persian. There is no doubt that the PA-Script is a serious bottleneck for non-Farsi speakers.

Finally, although this paper assumes some knowledge of Persian and the PA-Script, we hope that the fact the paper is written in English and also the large number of examples help others to find the paper useful.

The rest of this report is organized as follows. First the eFarsi alphabet is introduced and then some special cases that require more explanation and treatment are discussed. Throughout this paper the eFarsi text is printed in italics. Unfortunately, we have not been able to use the Persian-Arabic script, therefore, we use the transliterated names of the letters of the Persian-Arabic script for referring to them. To avoid confusion, these names are printed in bold italics.

This paper is essentially a proposal which consists of a number of conventions for writing Persian using the Latin alphabet. We have tried to number these proposals so in order to facilitate future references to each convention. Each convention is assigned a number and is printed according to a certain format. For example, here is the first convention numbered **EOC-1**:

EOC 1 (*eFarsi Orthographic Convention Format*)

eFarsi is a Latin-based alphabet and just like other Latin-based alphabets it uses the Arabic numerals.

1.1 Persian-Arabic Script

For many centuries a variation of the Arabic alphabet has been used by Iranians to write Persian. This alphabet includes all Arabic letters together with four Persian-specific letters.

Unfortunately, we do not have access to the PA-fonts, therefore we will list them by using their names. We will print them in bold throughout this paper in order to reduce the risk of confusion.

Alef, Be, Pe, Te, Se, Jim, Ce, He (Hotti), Xe, Dâl, Zâl, Re, Ze, Že, Sin, Šin, Sâd, Zâd, Tâ, Zâ, Eyn, Qeyn, Fe, Qâf, Kâf, Gâf, Lâm, Mim, Nun, Vâv, He (Havvaz), Ye, Hamze

The four letters, ***Pe, Ce, Že*** and ***Gâf*** are only used in Persian. The nine letters ***Se, He-ye-Hotti, Zâl, Sâd, Zâd, Tâ, Zâ, Eyn*** and ***Qâf*** are imported from Arabic.¹

¹Although ***Se, He-ye-Hotti, Zâl, Sâd, Zâd, Tâ, Zâ, Eyn*** and ***Qâf*** are imported from Arabic, they are used in writing some Persian words such as *Kiumars* and *Esfahân* whose s-sounds are written using the Arabic ***Se*** and ***Sâd*** respectively. Another interesting issue is whether the letter ***Zâl*** is an Arabic letter. Some argue that the letter or a similar letter has existed in Persian. Here are some poets expressing their views [12]:

*Ânân ke be Fârsi soxan mirânand
 Dar ma'raz e dâl, zâl râ nanšânand
 Mâ qabl e vey ar sâken e joz vây bovad
 Dâl ast vagarna zâl e mo'jam xânand - Xâje Nasir*
 or
*Dar zabân e Fârsi farq e miân e dâl o zâl
 Bâ to guyam zânke nazdik e afâzel mobham ast
 Piš az u dar lafz e mo'frad gar sahih e sâken ast
 Dâl bâšad varna bâqi jomle zâl e mo'jam ast - Sharafaddîn Ali Yazdi*

PA-Script also has nine more characters or accents that decorate a letter and modify its role and sound. These are: *Zir*, *Zebar*, *Piŝ*, *Sokun*, *Ye-ye-Kutâh*, *Taŝdid*, three sorts of so the called *Tanvin* characters that we shall call *Do Zir*, *Do Zebar* and *Do Piŝ*.

Some aspects of the traditional PA-Script can be characterized as follows:

- *Writing direction*: In PA-Script, with the exception of numbers, text is written from right to left. Numbers are written from left to right. However, the numerals used are not the Arabic numerals.
- *Vowels are implicit*: PA-Script lacks vowels. In principle the accents of the Arabic script (***Kasre***, ***Zamme***, ***Fathe***, ***Sokun***²) are the vowels of the writing system, but in practice these vowels are left out. For experienced people this is not a problem, but for novices it creates a major problem and also creates a bottleneck for the new learners.
- *Many to many relationship between letters and sounds*: The same sound, for example z-sound, is written by one of the letters *ze*, *zâl*, *zâd* and *zâ*³. Several different sounds are represented by the same letter, for example, the occurrence of the Persian-Arabic letter ***Vâv*** in the words: *to* (you), *ru* (on, above), *vali* (but), *rowŝan* (bright). In these words, *o*, *u*, *v* and *ow* are all used to transliterate various occurrences of ***Vâv***. There are similar cases in English, the letter s, for example, in the words: choose (z), vision (zh), mass (s), tension (sh).
- *Lack of standard code*: There is no widely used standard coding system for PA-Script.

1.2 General Characteristics of eFarsi

Here is a brief list of the general characteristics of the scheme we are proposing.

- eFarsi transliteration uses an extended Latin-alphabet and is written from left to right.
- Numbers are written in Latin using the so called Arabic numerals *1*, *2*, *3*, *4*, *5*, *6*, *7*, *8*, *9*, *0*. Numbers are also written from left to right.

²***Sokun*** is used to indicate non-vocal letters

³These letters are distinct Arabic letters with distinct sounds and the problem we are describing here only concerns Persian. Furthermore, their names are pronounced differently in Arabic.

- eFarsi alphabet is adequate for most dialects of Persian.⁴ This issue naturally deserves more discussion, however, in this report we will not treat the issue any further.
- eFarsi is primarily based on pronunciation⁵. In other words, whatever is written is pronounced and vice versa. Some exceptions are discussed in Section ??.
- Although the aim is to be able to write everything which is pronounced, we have tried to minimize the number of the characters in the alphabet, which has meant minor compromises.⁶
- Each letter or diphthong corresponds to one sound.
- Capitalization, punctuation and abbreviation rules are mostly similar to other Latin-based scripts.

2 eFarsi Alphabet

We begin by introducing the transliteration alphabet. This alphabet does not introduce any major changes to what people already use and is also very similar to other existing proposals. Introduction of a new alphabet is only a small part of the proposal we present here and most of the issues we cover in eFarsi scheme (such as rules for writing compound rules, capitalization and abbreviation) are not limited to a particular alphabet. eFarsi has 30 characters: \hat{a} , a , b , c , d , e , f , g , h , i , j , k , l , m , n , o , p , q , r , s , \hat{s} , t , u , v , w , x , y , z , \hat{z} , ' and two diphthongs: ow and ey .

This alphabet contains the following vowels: \hat{a} , a , e , i , o , u . The letters \hat{a} , i and u are long vowels and a , e and o are short.

EOC 2 (*The eFarsi Alphabet*)

The eFarsi alphabet consists of 30 letters: \hat{a} , a , b , c , d , e , f , g , h , i , j , k , l , m , n , o , p , q , r , s , \hat{s} , t , u , v , w , x , y , z , \hat{z} , '.

EOC 3 (*Letter Names*)

⁴The examples in the report are written in the main stream accent of Iran.

⁵Following Aristoteles' words: 'Writing is the mirror of speech.'

⁶The PA-letters *Qeyn* and *Qâf* are pronounced slightly differently, but this difference is only apparent in a small number of words. Therefore, we have decided to only designate one eFarsi-letter q to this sound.

The eFarsi letters and their names:

\hat{a} (\hat{A}), a (A), b (Be), c (Ce), d (De) e (E), f (Fe), g (Ge), h (He),
 i (I), j (Je), k (Ke), l (Le), m (Me), n (Ne), o (O), p (Pe), q (Qe),
 r (Re), s (Se), \hat{s} ($\hat{S}e$), t (Te), u (U), v (Ve), w (Ow), x (Xe), y
(Ye), z (Ze), \hat{z} ($\hat{Z}e$), ' (Ist)

The names of the eFarsi alphabet are selected by the author and have no relation to the Arabic or Persian names of the corresponding letters in the PA-alphabet. The pronunciation of the names follow the eFarsi pronunciation rules that will be explained later in this section.

EOC 4 (*eFarsi Diphthongs*)

eFarsi has two special diphthongs: *ow* and *ey*

EOC 5 (*eFarsi Vowels*)

eFarsi has 6 vowels: \hat{a} , a , e , i , o and u . The vowels \hat{a} , i and u are long whereas a , e and o are short.

The rest of this section provides brief explanations of each letter and diphthong. In connection with the description of each letter we provide some examples as further clarification. In some cases, where the pronunciation of a letter is not similar to the corresponding English letter, we provide some English words written in eFarsi in order to provide the readers further information about the correct sound associated with the eFarsi-letter.

\hat{a} (\hat{A})

- transliteration of certain occurrences of the Persian-Arabic letter **Alef** such as *bâd* (wind), *bâdâm* (almond), *jodâ* (separate)
- transliteration of the **Alef e maddâh** (the Arabic name of the letter) or **\hat{A} -ye-bâ-kolâh** (the Persian name of the letter), for example, *âb* (water), *âdam* (Adam, human), *âsmân* (sky)
- transliteration of the **Alef e bâlâ** or **Alef e maqsure**⁷, for example, in *Allâh* (God), *lâken* (but), *Musâ* (Moses), *Esmâ'îl*, *Isâ* (Jesus). This transliteration rule may be overruled by poets for other considerations, for example, in the poem *Isi e Maryam be kuhî migorixt*, *Isi* is used instead of *Isâ* allowing the poet to create better rythm in the poem by exploiting the Arabic spelling of the word (which ends with an i-sound).

⁷In Arabic script this takes the form of a small **Alef** placed on top of a final **Ye**.

- â is pronounced as the *a* in 'bar', 'tall'
- some English words written in eFarsi: *tâl* (tall), *gelâs* (glass), *Amerikâ* (America), *Irân* (Iran), *âl* (all)

a (A)

- transliteration of the vowel **Zebar** or **Fathe**, for example, *bad* (bad), *sar* (head), *dar* (door), *pesar* (boy, son). **Zebar** is rarely written in PA-Script.
- transliteration of certain occurrences of **Alef** appearing at the beginning of a word⁸ with an implicit or explicit **Zebar** (**Fathe** in Arabic), for example, *abr* (cloud), *asb* (horse), *andišeh* (thought), *Allâh* (God), *akbar* (great), *Aflâtun* (Plato)
- transliteration of some occurrences of **Eyn** appearing at the beginning of the word with an implicit or explicit **Zebar**, for example, *Ali* (Ali), *aks* (photo, image, negation, negative), *arabi* (Arabic), *alâqe* (interest, passion)
- *a* is pronounced as the English 'a' in 'and', 'bad' but never as the 'a' in 'ball', 'far' or 'take'
- some English words written in eFarsi: *kan* (can), *abzorb* (absorb), *asowsieyt* (associate)

b (Be)

- transliteration of the letter **Be**, for example, *âb* (water), *baxš* (partition, division, part), *bâbâ* (dad, father), *tabdil* (exchange), *bâdâm* (almond), *mohabbat* (kindness), *babr* (tiger)
- pronounced exactly like the *b* in English

c (Ce)

- transliteration of the Persian letter **Ce**, for example, *cub* (wood), *bacce* (child), *cerâ* (why), *kucak* (small), *cekke* (drop), *casb* (glue), *barcasb* (label), *murce* (ant), *parce* (cloth, textile), *cahâr* (four)

⁸An **Alef** that appears at the beginning of a word usually denotes a *Hamze* (or a glottal stop). This glottal stop usually functions as the bearer of a vowel sign (**Zebar**, **Zir**, **Pîš**) which is seldom written and it is the reader that has to determine the sign.

- pronounced exactly like the *ch* in English
- some English words written in eFarsi: *cans* (chance), *lânc* (lunch)

d (De)

- transliteration of the letter **Dâl**, for example, *dar* (door), *mâdar* (mother), *pedar* (father), *barâdar* (brother), *radd* (trace), *hadd* (limit), *medâd* (pencil)
- pronounced exactly like the English *d*

e (E)

- transliteration of certain occurrences of **Alef** at the beginning of a word with an implicit or explicit *Zir* vowel, such as, *Ebrâhim* (Abraham), *ensân* (human), *Eslâm* (Islam)
- transliteration of the vowel **Zir** (**Kasra** in Arabic), for example, *deraxt* (tree), *del* (heart), *pedar* (father), *hezâr* (thousand), *yek* (one)
- transliteration of certain occurrences of the letter **Eyn** in association with an implicit or explicit **Zir** vowel, for example, *ešq* (love), *elm* (science)⁹
- When *e* occurs at the beginning of a word or after a vowel, it is pronounced as the initial 'e' in the English words 'end', 'ever', 'England'. The occurrences of *e* in the middle or the end of a word are pronounced the same way and the pronunciation is similar to the 'e' in the English words 'pen', 'flower' and 'net'. The occurrences of *e* at the end of eFarsi words are pronounced and the pronunciation is similar to the 'e' in the English words 'water', 'regard'. So if *elite* were to be pronounced as a Persian word it would have been pronounced as *é-li-té*
- some English words written in eFarsi: *sentens* (sentence), *hed* (head), *leter* (letter)

f (Fe)

- transliteration of the letter **Fe**, for example, *farâvân* (many, much), *barf* (snow), *raftâr* (behaviour), *daftar* (office, book), *raft* (went), *fahmidan* (to understand)

⁹Persian word for science is *dâneš*.

- pronounced exactly like the 'f' in English

g (Ge)

- *g* is used for transliteration of the Persian letter **Gâf**, for example, *agar* (if), *dânešgâh* (university), *gâhi* (sometimes), *gonâh* (sin), *marg* (death), *gâv* (cow, ox), *giti* (world), *rag* (vein), *barg* (leaf), *gorg* (wolf), *gerowgân* (hostage), *sahmgîn* (frightening), *golestân* (flower garden)
- pronounced as the English 'g' in 'good', 'fog', but never as the 'g' in 'George' or 'gene'.

h (He)

- *h* is used for transliteration of the letters **He-ye-havvaz** and **He-ye-hotti**, for example, *Hamid* (a male name), *har* (any), *behtar* (better), *rahbar* (leader), *bahâr* (spring), *mâh* (moon), *kahkešân* (galaxy), *haft* (seven), *hašt* (eight), *noh* (nine), *Nuh* (Noah).
- pronounced as the 'h' in the English words 'hat', 'hot'.¹⁰

i (I)

- transliteration of the PA-letter **Ye** when used as a vowel, for example, *sib* (apple), *riz* (tiny), *niyat* (intention), *pâiz* (autumn), *šiš* (six), *bini* (nose), *šir* (lion, tap, milk), *sefid* (white), *siâh* (black), *sir* (garlic, full - not-hungry)
- With the exception of a few instances, most of the I-sounds in Persian are long (similar to -ee- in 'been')¹¹. Some exceptions (where the *i* is pronounced as in 'bit') are *šiš* (six), *hijdah* (eighteen), *milyon* (million), *šiš o biš* (six and five in backgammon, *biš* is the Turkish for five), *mihmân*¹² (guest), *binâi* (vision).
- some English words written in eFarsi: *bin* (bean), *bin* (been), *bin* (bin), *sin* (sin), *sin* (seen), *Irân* (Iran).

¹⁰In Arabic, the letters *He-ye-havvaz* and *He-ye-hotti* are pronounced differently, but their Persian pronunciations are the same.

¹¹The motivation for not choosing to have another letter or diphthong for the short *i*-sound is that there are so few words with the short I-sound.

¹²The *i* in *mihmân* is also pronounced as a long *i*. When the *i* is short it is also written as *mehmân* simply because the *e* seems to be a good approximation for a short *i* in some cases.

j (Je)

- transliteration of the letter **Jim**, for example, *jân* (spirit), *xarj* (cost, expenditure), *jâ* (place), *injâ* (here), *ânjâ* (there), *juybâr* (water stream), *jâmedân*¹³ (suitcase)
- pronounced as the 'j' in 'John', 'major'
- some English words written in eFarsi: *Jorj* (George), *Jân* (John), *jus* (juice)

k (Ke)

- transliteration of the letter **Kâf**, for example, *kâr* (work), *kârgar* (worker), *namak* (salt), *kušêš* (try), *pâk* (clean, pure), *dardnâk* (painful), *kârvân* (caravan), *dâneškade* (university department), *meykade* (bar/place serving wine and other alcoholic drinks), *âtaškade* ('place of fire', holy places for Zarathustrians)
- pronounced as the *k* in 'book', 'kitten'

l (Le)

- transliteration of the letter **Lâm**, such as, *lâle* (tulip), *pol* (bridge), *kalle* (head), *lop* (cheek), *pašm* (wool), *del* (heart), *deldâr* (person who is in love with someone), *Landan* (London)
- pronounced as the 'l' in 'love', 'label'

m (Me)

- transliteration of the PA-letter **Mim**, for example, *miz* (table), *tamiz* (clean), *âsmân* (sky), *caman* (grass), *zamân* (time), *mâmân* (mom, mother), *mâdar* (mother), *xeradmand* (intelligent person with common sense), *bâmdâd* (morning), *meh* (fog, May), *mâh* (moon), *meqnâtis* (magnetism), *miravam* (I am going), *hamsâye* (neighbour), *mard* (man)
- pronounced as the 'm' in English

¹³*jâmedân* is not used anymore, *camedân* is more common.

n (Ne)

- transliteration of the letter **Nun**, for example, *Tehrân* (Teheran), *nazm* (order), *namakdân* (salt jar), *kamân* (bow), *tofang* (gun), *boland* (tall), *šen* (sand), *zan* (woman), *nik* (good), *angošt* (finger, toe), *nân* (bread), *band* (string)
- *n* is also the transliteration of some other Arabic-manifestations of the n-sound, so called **Tanvin**, which occurs at the end of words.¹⁴ For example, *fe'lan* (for the time being), *ehemâlan* (probably), *mozâfon elayh* (that to which is added)
- pronounced as the English 'n' in, for example, 'number', 'fan'

o (O)

- transliteration of the vowel **Piš** (**Zamme** in Arabic), for example, *Morq* (bird), *mohemm* (important), *o* (short form of *va* (and)), *otu* (iron - for ironing clothes)
- transliteration of some occurrences of the letter **Alef** equipped with an explicit or implicit **Piš**, for example, *Otriš* (Austria), *ordak* (duck), *otobus* (Bus), *otomâtik* (automatic), *Orupâ* (Europe), *otu* (iron - for ironing)

¹⁴Some Arabic words end with an n-sound with no explicit **Nun** at the end of the word. When these letters are made explicit, they are written with an accent similar to a double quote which heads an **Alef** unless the Arabic word ends with a **Tâ ye Zâ'ed** (redundant **Tâ** or silent **He ye Havvaz**) or a **Hamze** in which case the **Tanvin** is written without an **Alef**. In any case, as far as the transliteration is concerned we have a simpler situation and we only have the following three cases:

- **Fathatan** (or **Tanvin e Fathe** or **Do Zebâr**) is transliterated as *-an*, for example, *masalan* (for example), *eynan* (exactly, obviously), *aslan* (originally, inherently), *abadan* (never, not at all), *ettefâqan* (As it happens, coincidentally), *qat'an* (certainly), *amdan* (on purpose), *haqiqatan* (really), *âdatan* (as of habit), *ebtedâ'an* (initially)
- **Dammatan** (or **Tanvin e Zamme Do Piš**) is transliterated as *-on*, for example, the Arabic *mozâfon elayh* (that to which is added)
- **Kasratan** (or **Tanvin e Kasre** or **Do Zir**) is transliterated as *-en*, for example, *ahaden*, *vâleden*

Tanvin only appears in Arabic words. Persian words that contain **Tanvin** are incorrect, for example, some people occasionally use these words but they are incorrect: *zabânan* (expressing something orally), *nâcâran* (out of necessity). The correct alternatives for *zabânan* and *nâcâran* are *bâ zabân* and *az nâcâri*.

- transliteration of some instances of the letter **Vâv**, for example, *do* (two), *vidio* (Video), *xod* (self), *xošgel* (beautiful)
- pronounced as the 'o' in the English words 'normal', 'port'.

p (Pe)

- transliteration of the Persian letter **Pe**, for example, *pul* (money), *tup* (ball)
- pronounced as the English letter 'p'.

q (Qe)

- transliteration of the Arabic letter **Qâf** and the PA-letter **Qeyn**¹⁵, for example, *gamgin* (sad), *Qom* (Iranian holy city of Qom), *barq* (electricity)
- pronounced almost as the French letter 'r' in 'rue', but often the pronunciation is heavier than in French somewhat close to the Scottish 'ch' in, for example, 'luch'. There are no corresponding sounds in the modern English.

r (Re)

- transliteration of the letter **Re**, for example, *Rom* (Rome), *narm* (soft, flexible), *xar* (donkey). Here are some more examples, all of which are written the same way in Arabic script: *korre* (foal), *Kore* (Korea), *kore* (sphere, globe), *kare* (butter), *karh* (reluctance).
- pronounced just like the English letter 'r'.

s (Se)

- transliteration of the letters, **Se**, **Sin** and **Sâd**, for example, *Sorayyâ* (Arabic name for women), *serke* (vinegar), *sabr* (patience), *tars* (fear)
- pronounced as the 's' in 'safe', 'parse'

¹⁵ **Qâf** and the PA-letter **Qeyn** are pronounced differently in Arabic, but in Persian the pronunciations are in general very similar. In Persian and some Arabic dialects, many occurrences of **Qâf** are pronounced as **Qeyn**. This is the motivation of transliterating both to *q* in eFarsi.

š (Še)

- transliteration of the letter **Šin**, for example, *šab* (night), *kešvar* (country), *âš* (soup)
- pronounced as the 'sh' in 'shuttle', 'push'

t (Te)

- transliteration of the letters **Te**, **Tâ** and the so called **Te-ye-marbuta**, for example, *tâq* (ceiling), *kot* (coat, jacket), *hattâ* (even), *dâ'erat alma'âref* (encyclopedia), *hotel* (hotel)
- pronounced as the 't' in 'tea', 'root', 'hotel'

u (U)

- transliteration of some occurrences of the letter **Vâv**, for example, *u* (he, she), *bu* (smell), *ruz* (day), *zud* (early), *abru* (eyebrow), *Urânus* (Uranus), *ut* (August)¹⁶.
- pronounced as the 'oo' in 'tool', 'cool'

v (Ve)

- transliteration of some occurrences of the letter **Vâv**, for example, *va* (and), *miravam* (I am going), *vârune* (upside down)
- pronounced as the 'v' in 'victory', 'traverse'

x (Xe)

- transliteration of the letter **Xe**, for example, *xâk* (soil), *xub* (good, nice)
- pronounced as the 'ch' in German 'achtung' or the 'ch' in Dutch 'acht' or the 'g' in the Dutch word 'geweldig' or the diphthong 'kh' in 'khaki', 'Khamanei' and 'Khatami'.

¹⁶Very few Persian words start with the letter *u*

y (Yâ)

- transliteration of the letter PA-letter **Ye** in situations where it is used as a consonant, for example, *yax* (ice), *niyat* (intention), *Xayyâm* (Persian poet and mathematician)
- pronounced as the 'y' in 'young', 'yes', 'guy'

z (Ze)

- transliteration of the letters **Ze**, **Zâl**, **Zâd** and **Zâ**, for example, *zar* (gold), *ruz* (day), *gâz* (gas), *zedd* (opposite, against), *zarf* (container, holder), *lahze* (moment)
- pronounced as the 'z' in 'zero', 'buzz'

ẑ (Zê)

- transliteration of the Persian letter **Zê**, for example, *zâle* (dew), *zârf* (depth), *dež* (castle), *Senadež* (original name of Sanandaj)
- pronounced as the French 'j'. In English, some occurrences of the letters 'g' and 's' are also pronounced like the Persian *ẑ*, for example, the 's' in 'vision' or 'pleasure' and the second 'g' in 'Garage' (in American pronunciation).
- some English words written in eFarsi: *vižen* (vision), *peležer* (pleasure), *režim* (regime)

' (Ist)

- transliteration of the letter **Eyn**. **Eyn** only appears in Arabic words. For example, *mo'allem*¹⁷ (teacher), *bo'd* (dimension)
- transliteration of the **Hamze**. In Persian, **Hamze** only appears in a medial position and indicates that there is a change of vowels. For example, *pâ'iz* (autumn), *miâ'i* (you come), *migu'im* (we say). In some cases, **Hamze** in Persian words has been replaced by a **ye**. For *pâyiz* instead of *pâ'iz*, *miguyim* instead of *migu'im*, *rayis* instead of *ra'is*.

¹⁷The Persian word for the Arabic *mo'allem* is *Âmuzgâr*.

- pronounced in the throat, just like the way the second 't' in 'that' or 'tt' in 'bottle' is pronounced in some English accents (Cockney - in London or Mancunian - in Manchester). The sound generated by *Ist* is what linguists call a *glottal stop* which appears in many languages. Compare, for example, the sounds of the first and the second *a* in the English word *Adam*. The first *a* in Adam starts with a glottal stop.

ey (Ey)

- transliteration of certain occurrences of the Persian-Arabic letter **Ye**, for example, *peykar* (body), *peykân* (arrow), *peyâpey* (repeatedly), *key* (when), *dey* (10th month in the Iranian calendar, yesterday)
- pronounced as the *a* in the English 'take', *ei* in the English word 'eight' or the *ai* in 'Bahrain', but never as *ey* in the English word 'key'.
- some English words written in eFarsi: *teyk* (take), *eyt* (eight), *Bahreyn* (Bahrain), *neyber* (neighbour)

ow (Ow)

- transliteration of certain occurrences of the letter **Vâv**, for example, *mowlâ* (sir), *nowruz* (new year's day), *partow* (ray), *peyrow* (follower). *w* does not occur alone in eFarsi.
- pronounced as the *ow* in *bow*, *snow*. Sometimes, in spoken Persian, if the letter following *w* is a vowel, then *w* is pronounced like a *v*, for example, *peyrowân* (followers)

In section 3.10, a few more diphthongs are presented for practical reasons.

3 eFarsi Conventions

3.1 Writing Words As Pronounced

eFarsi writing conventions are mainly based on pronunciation. The goal is to keep the spelling as close to the pronunciation as possible. For example, the eFarsi word *xâhar* is written as *xvâhr* in the PA-Script. But since the letter *v* is not pronounced in modern Persian, it does not appear in the eFarsi spelling.

EOC 6 (*eFarsi is Based on Pronunciation*)

eFarsi writing is generally based on pronunciation. That which is pronounced is written and vice versa.

There are a number of exceptions to EOC-6, when the letters *n* and *b* appear immediately after each other, the *n* is usually pronounced as *m*, a phenomenon that appears in other languages too. For example, *panbe* (cotton) is pronounced as *pam-be*, *šanbe* (Saturday) as *šam-be*, *donbe* (fat of sheep) as *dom-be*. In these cases, we have decided not to transliterate the *nb* as *mb* because we think the *m*-sound is a surface phenomenon which is simply the result of trying to pronounce *n* and *b* after each other. Anyhow, there are exceptions to this exception; for example, words like *dom* (tail) and *som* (hoof) that are established Persian words that originate from *donb* (tail) and *sonb* (hoof).

3.1.1 Some Arabic-Related Issues

Whether we like it or not, Arabic is an important ingredient of Persian. Persian texts often contain Arabic words and terms and, most important of all, many great Iranian writers, poets, mathematician and scientists have Arabic names. It is therefore, important to extend the conventions of eFarsi to include some issues that are specific to Arabic.

The Arabic Definite Article *Al*

The Arabic definite article *al*, which is used as a prefix, is not always pronounced the same way. For example,

1. in *almo'menun* (the believers) or *alqamar* (the moon) the pronunciation of *al* is as written; whereas,
2. in *alra's* (the head) or *alšams* (the sun), the *l* in *al* inherits the pronunciation of the first letter of the prefixed word. That is, *alra's* is pronounced as *arra's* and *alšams* as *aššams*.

The first case applies to the so called *Alhorofalqamariya* ('the moon letters') and the second to the *Alhorofaššamsiya* ('the sun letters'). Here is a list of *Alhorofalqamariya*:

Alef, Be, Jim, He (Hotti), Xe, Eyn, Qeyn, Fe, Qâf, Kâf, Mim, Vâv, He (Havvaz), Ye, Hamze

And the rest of the letters, *Alhorofaššamsiya*, are:¹⁸

***Te, Se, Dâl, Zâl, Re, Ze, Sin, Šin, Sâd, Zâd, Tâ, Zâ, Lâm,
Nun***

In other words, there is a choice between two different conventions for Arabic words starting with *Alhorofaššamsiya* that are prefixed with the Arabic definite article *al*:

1. treat *Alhorofalsamsiya* and *Alhorofalqamariya* the same way
2. let the letter *l* of the Arabic definite article *al*, which prefixes *Alhorofaššamsiya*, be transformed to the 'sun letter' following it

The first case would mean uniformity in writing the article *al* but breaking the convention EOC-6. It would also mean that the speaker or the reader would have to be aware of 'sun letters' in order to be able to pronounce words correctly.

The second case has the advantage of respecting the EOC-6 convention but undermining the uniformity of writing the definite article. The first case creates an exception at speech time and the second an exception at writing time. We have chosen the second alternative.

Al and Hamzatolvasl

In Arabic certain occurrences of ***Hamze***, so called ***Hamzatolvasl***, are dropped in certain situations. The *a* of the definite article *al* which is a ***Hamze*** is dropped in certain contexts. For example, the compound word *Malekoššo'arâ* (king of the poets) is the result of combining *Malek*, *o* and *alšo'arâ*. *a* of *al* has been dropped since it is treated as ***Hamzatolvasl*** and the *l* of *al* has been transformed to *š* since *š* is a 'sun letter'. Here are some more examples: *yadollâh* [*yad+o+Allâh*] (God's hand), *abolmoslem* [*abu+almoslem*] where the *a* of *al* has been transformed to *o* which is the final letter¹⁹ of the preceding word.

EOC 7 (*Arabic Al*)

¹⁸An easy way to remember these letters (at least in eFarsi) is to think of them as the consonants of the expression *šenel zardast*

¹⁹This is not true, what has really happened is that the *u* of *abu* which is a long vowel has, in this context, been transformed to its shorter version *o*. Similarly, *â* and *i* can be shortened to *a* and *e* respectively.

The Arabic definite article *al* is written as *al* when the prefixed word starts with a 'moon letter'. If the word starts with a 'sun letter', then the *l* of the article assumes the pronunciation, and thereby, the spelling of that 'sun letter'. In some other situations the *a* of *al* may be dropped.

3.2 *Ist*

Ist, the glottal stop, is denoted by a single quote or apostrophe (') and used for transliterating certain occurrences of the PA-letters *Eyn* and *Hamze* and is usually used in words that are originally Arabic.

The glottal stop phenomenon occurs in many languages including many dialects of English²⁰ but no written form is associated with it in the Latin alphabet. In Persian, making the *Ist* explicit is not essential in many cases, but it is sometimes useful for ensuring the correct pronunciation. For example, for distinguishing the pronunciation of the Persian word *bad* (bad) and the Arabic word *ba'd* (after, later).

In speech, *Ist* appears at the beginning, middle or the end of a word. But in writing it is most important to include the stops that appear in the middle of the word. Even the stops appearing at the end of the words may be useful to include in the written text. Here are some examples of *Ist* appearing in the middle or end of a word: *mas'ale* (problem), *so'âl* (question), *âoru'* (start or beginning), *tolu'* (sunrise). We feel that making *Ist* explicit at the beginning of a word is not necessary since the pronunciation of the vowels appearing at the beginning of a word enforce the same effect. For example, pronunciation of *'âb* is not different from *âb* (water).

Finally, *Ist* also occurs between two vowels. For example, *jâme'e* (society), *fa'âl* (active), *so'âl* (question).²¹ See Section 3.11.3 for further discussion of issues related to *Ist*.

²⁰Glottal stops are common in English, for example, when pronouncing the expression 'uh-uh' meaning 'no' [9]. In speech, each of the vowels used for this expression are preceded by a glottal stop which is articulated by a momentary stop of airflow at the glottis. In some English dialects, in words such as *better* or *bottle*, the -tt- sound usually is accompanied by a stop. A similar situation occurs in words such as *button* or *written* in most American English dialects [9]. Glottal stop also occurs at the beginning of a lot of English words that begin with a vowel. For example, notice how the first letters of the following words are pronounced: *epsilon*, *end*, *apple*, *and*, *odd*, *Iran*. Now you may close your eyes and say these words once more and try to be aware of what is happening in your throat when pronouncing the initial vowel.

²¹Here are some more examples: *mo'men* (man of god), *ta'min* (been provided for), *masâ'el* (problems), *ajzâ'* (parts), *xala'* (vacuum), *me'mâr* (architect), *te'dâd* (number), *qâte'* (decisive), *sâ'at* (watch, clock).

EOC 8 (*Ist*)

If *Ist* occurs in the middle or at the end of a word, it is written explicitly.

3.3 No Silent *h*

Lack of vowels in the PA-Script has created some strange phenomenon in the writing system. One of these is addition of the PA-Letter **He** to the end of a word simply for the sake of indicating that the word ends with a vowel²². The **He** itself is not pronounced. The silent *h* can only appear at the end of a word. In traditional transliteration schemes for Persian, the silent *h* is usually included for some reason. For example, in *padideh* (phenomenon), *Zohreh* (Venus), *šâneh* (comb) and *parandeh* (bird).

In eFarsi the silent *h* is dropped since it is redundant. This is also true of the transliteration schemes Eurofarsi [3] and UniPers [4]. The above examples, are written as *padide*, *Zohre*, *šâne* and *parande*.

3.4 *Tašdid*

Tašdid in Persian and Arabic is a sign (similar to a small *w*) which is (supposed to be) placed on a consonant²³ to indicate that the consonant should be pronounced twice with a short pause between the pronunciation of the two. The pronunciation of the first instance is not vocalised but the second one is succeeded by a vowel. Most transliteration schemes indicate this by writing the consonant twice. eFarsi follows the same principle.

In English, for example, many words include double consecutive consonants such as *little* and *connect* but there is a major difference between Persian and English in that, in Persian, both occurrences are pronounced. Here are some example words: *matte* (drill) - pronounced as matt-te, *bacce* (child) - pronounced as batch-che, *fann* (technique) - pronounced as fan-n, *jâdde* (road) - pronounced as jâd-de, *tavallod* (birth) - pronounced as taval-lod.

²²In Tehrani Persian the vowel is usually *e* whereas in other dialects it would be *a*

²³In Persian *Tašdid* is only applied to the letters **Pe**, **Ce**, **Re** and **Lâm**, for example, *tappe* (hill), *bacce* (child. kid), *arre* (saw), *galle* (flock). In Arabic all letters except **Alef** can assume a *Tašdid*. Another interesting fact is that the second letter of most two-letter Arabic words assumes a *Tašdid*, for example, *omm* (mother), *xatt* (writing, calligraphy), *serr* (secret), *hadd* (limit). The two occurrences of the consonant in the singular word even appear in the its plural form, for example, *asrâr* (plural of *serr*), *hodud* (plural of *hadd*). This is not true of Arabic two-letter words without *Tašdid*, for example, the Arabic word *ab* (father) and its plural *abâ*?

EOC 9 (*Tašdid*)

In eFarsi, a PA-letter with a *Tašdid* is transliterated to two subsequent occurrences of the eFarsi equivalent of that PA-letter.

3.4.1 *Tašdid* on the Persian-Arabic letter *Ye*

In eFarsi, the PA-letter *Ye* is either transliterated as *i* or *y*; this depends on whether it has the role of a vowel²⁴ or a consonant.

If there is a *Tašdid* on *Ye*, then it is transliterated to either *yy* or *iy*. In the *iy*-transliteration, which is more common, the first *Ye* is treated as a vowel and the second one as a consonant. For example, *niyat* (intention), *ahammiyat* (importance), *movaffaqiyat* (success), *amaliyât* (operations). In the *yy*-transliteration, both *Ye*:s are treated as consonants, for example: *xayyât* (tailor), *Xayyâm* (Famous Iranian poet and mathematician).

EOC 10 (*Exception to EOC-9*)

Tašdid on *Ye* is either transliterated as *iy* or *yy*.

3.4.2 *Tašdid* on the final letter

Some words (usually loan-words from Arabic,) have a *Tašdid* on the final letter; for example, *fann* (technology), *hadd* (limit), *radd* (trace), *zedd* (opposite, against), *sadd* (dam, barrier)²⁵. The pronunciation of the second occurrence of the letter is more apparent only when the word is succeeded by a vowel. For example, the pronunciation of *sadd* in *sadd e Karaj* (Karaj dam) and *sadd kardan* (to block). In the latter case the second *d* is not pronounced.

3.5 Ezâfe

Ezâfe (*e* or *ye*) is a prepositional morpheme used to indicate a relationship between two words. In PA-script it is not always written. In eFarsi it is transliterated according to one of the following formats:

²⁴Here are some English words where the letters *i* and *y* have the role of a vowel, we write them in eFarsi in parantheses: 'honey' (*hâni*), 'in' (*in*), 'idiom' (*idiom*), and as consonants: 'yacht' (*yât*), 'yellow' (*yelow*), 'try' (*terây*).

²⁵In the PA-Script, these words consist of two letters only with a *Tašdid* on the second letter. When these words are used for deriving other words according to the Arabic grammar, the double occurrence of the second letter becomes explicit. For example, note the two occurrences of *n* in *fonun* (plural of *fann*), and the two occurrences of *d* in *mahdud* (limited - derived from *hadd*)

1. <mozâf> e <mozâfon elayh>
2. <mozâf> ye <mozâfon elayh>

EOC 11 (*Ezâfe*)

Ezâfe is transliterated as *e* or *ye*. If *Mozâf* ends with an *â*, *o*, *u* or *e* then it is followed by a *ye* instead of *e*.

Ezâfe is used for different purposes. Sometimes it functions as the English *'s* indicating possession and sometimes it has the same role as the French *de*²⁶ or the English *of*²⁷. The words preceding and succeeding the *Ezâfe* are called *Mozâf* and *Mozâfon Elayh* respectively. For example, *ketâb e man* (my book), here, *ketâb* (book) is the *Mozâf* and *man* (I) is the *Mozâfon Elayh*. The *Ezâfe* can be used to create arbitrarily complicated expressions, for example, *ketâb e târix e moâser e jahân e man*²⁸ (my contemporary history of the world book).

In the traditional PA-Script, *Ezâfe* seldom appears in written text. Only in certain cases where it has the *ye*-form (case 2 above) the letter **ye** is added to the word, for example, when writing the eFarsi expression *pahlu ye to* in PA-Script *pahlu* is followed by a **ye**. In eFarsi, however, it is written whenever it is pronounced.

EOC 12 (*Writing Ezâfe*)

Ezâfe is always written.

Because Ezâfe declares some sort of correspondence between two entities, we have chosen to write it separately rather than concatenating it to the end of the *Mozâf*²⁹ which is the case in some other schemes ([?]). Furthermore, writing it separately facilitates text analysis to some extent.

EOC 13 (*Writing Ezâfe*)

Ezâfe is normally written as a separate word.

²⁶'Rue de Lafayette' or 'Jean de Arc'

²⁷'Moment of weakness' or 'Princess of Wales'

²⁸In *ketâb e târix e moâser e jahân e man* the words can be grouped as

[[ketâb e [[târix e moâser] e jahân] e man]

²⁹That would be unfair to the *Mozâfon Elayh*.

Ezâfe appears in many different contexts (See Appendix B): possessivity (*ketâb e man* (my book), specialization of a concept (*saqf e xâne* (Ceiling of a house)), type specification (*samâvar e noqre* (silver samovar)), comparison (*mahd e Zamin* (mother Earth)), metaphorical (*dast e ruzgâr* (hand of time)), connecting noun and its adjective (*âb e zolâl* (clear water)). Here are some more examples:³⁰

ruz e âftâbi (a sunny day)
ruz (day), *âftâb* (sun), *âftabi* (sunny)

ketâb e man (my book)
ketâb (book), *man* (I)

miz e motâle'e (study table)
miz (table), *motâle'e* (study)

rang e sorx (red color)
rang (color), *sorx* (red)

sedâ ye boland (loud voice)
sedâ (voice/sound), *boland* (loud)

ru ye miz (on the table, above the table)
ru (on, above), *miz* (table)

havâ ye xub (good air, nice weather)
havâ (air), *xub* (good, well, nice)

Some words are always succeeded by an *Ezâfe* and for this reason it is more appropriate to write the word and the *Ezâfe* as one word. For example, the word *barây* (for) and *bedun* (without) are always succeeded by an *Ezâfe* therefore we suggest that they should be written as *baraye* and *bedune* respectively.

EOC 14 (*Ezâfe, Exceptions to EOC-13*)

Ezâfe which succeeds *barây* (for) and *bedun* (without) is glued to these words which are subsequently written as *baraye* and *bedune*.

³⁰Even more examples: *cây e iran* (Iranian tea), *ra'y e mardom* (people's vote, the wish of the people); *baraye to* (for you), *ruz e âftâbi* (sunny day), *qorub e xorâid* (sunset), *joz' e kuchak* (small part), *farâs e dast bâft* (hand-made carpet), *kašti e Nuh* (Noah's ship), *xâne ye mâ* (our house), *kâdo ye tavallod* (birthday present), *jangju ye dalir* (brave warrior), *ketâb e Rîyâzi e Novin e man* (my Modern Mathematics book)

Sometimes when Ezâfe is used to connect a noun and an adjective, one can, in principle, switch the position of the noun and the adjective and thereby do without the Ezâfe. For example, *ruz e âftâbi* (sunny day) which can be rephrased as *âftâbi ruz* without change of meaning. This construction is quite common in some dialects of Persian and also in Persian poesi. In section 3.6.2, some further issues related to Ezâfe are discussed.

3.6 Compound Words

Compound words or terms are formed by joining two or more words or adding prefixes or suffixes to words. Such words are numerous in Persian.

3.6.1 Constituents of Compound Words

In this section we provide a list of word-types that are usually used for creating compound words.

1. Two nouns, for example, *kârxâne* [*kâr-xâne*] (factory), *šekarâb* [*šekar-âb*] (sweet drink made with suger and water), *Xalij e Fârs* (Persian Gulf), *âb o havâ* (climate), *cašm be râh* [*cašm* (eye), *be* (to, on), *râh* (road)] (state of waiting), *mâdaršowhar* [*mâdar* (mother) *šowhar* (husband)] (husband's mother), *sangdel* [*sang* (stone), *del* (heart)] (cruel), *sarâsar* (everywhere, whole)
2. Two verbs or verb roots, for example, *kešâkeš* [*keš-â-keš*] (struggle), *hast o nist* (belongings, 'that which exists and that which not'), *bud o nabud* (existence), *âmad nayâmad* (for example, *âmadnayâmad dâre* would mean: it may or may not 'stick'.)
3. A noun and an adjective *Sefidrud* [*Sefid-rud*] (Sefidrud - 'white river'), *Siâhkuh* [*siâh-kuh*] (Siâhkuh - 'black mountain'), *Nowruz* (Persian New Year - 'new day'), *xošlebâs* [*xoš* (nice), *lebâs* (clothes)] (well-dressed, nicely dressed), *zibâru* [*zibâ* (beautiful, pretty), *ru* (face)] (pretty face), *kamru* [*kam* (little), *ru* (face)] (shy), *kamzur* (weak), *porzur* (strong)
4. A noun and some form of a verb, for example, *pâdow* [*pâ* (foot), *dow* (Present Stem of *dauidan* (to run))] (springpojke), *sarafrâz* [*sar* (head), *afrâz* from *afrâstan*] (honoured), *mâdarxânde*, *darmânde* [*dar* (in), *mânde* is from *mândan* (to remain, to stay)] (hopeless)

5. A numeral and a noun, for example, *cârpâ* [*câr*³¹ (four) *pâ* (foot, leg)] (four-legged)
6. An adverb and a verb, for example, *pişrow* [*piş* (forward), *row* from *raftan*] (pioneer)
7. An adjective and a verb, for example, *zende-bâd* [*zende-bâd*] (long live), *şâdbâş* [*şâd-bâş*] (congratulation), *nowâmuz* [*now* (new) *âmuz* from *amuztan*] (new learner), *tondrow* [*tond* (quick, fast) *row* (walker, mover) from *raftan* (to go)] (fast mover, extremist)
8. An infinitive and a verbal noun connected by *o* the short form of the connective *va* (and), for example, *xordoxâb* [*xord-o-xâb*] (eating and sleeping)
9. Two rhyming words³² where the second one is constructed by replacing the first letter of the first word with *m* (and sometimes with *p*). The second word is in principle meaningless but could be taken as meaning 'and the like', for example, *pul-mul* (money), *catr-matr* (umbrella), *qâti-pâti* (mixed, without order)
10. Repetition of a noun, adjective or an adverb may in some cases be used to amplify the semantics of a word, for example, *nam nam* (in fine drops)³³, *zâr-zâr* (loud and bitter - when crying), *daste daste* (in bundles, in groups)
11. A word and a preposition, for example, *bikas* (someone with no friends or relatives), *barqarâr* (established), *navâred* (novice - in a negative sense)
12. A prefix and a word *begu* [*be-gu*] (say), *barnâme* [*bar-nâme*] (program), *pişraft* [*piş-raft*] (progress), *ham-kâr* [*ham* (like, same), *kâr* (work)] (co-worker), *nâ-dân* (ignorant)
13. A word and a suffix, for example, *lâlezâr* [*lâle-zâr*] (tulip garden), *kârgar* [*kâr-gar*] (worker), *behtar* [*beh* (good)] (better), *behtarin* (best), *dâneşmand* [*dâneş* (knowledge)] scientist,)
14. Adding an infix, for example, *keş-â-keş* (struggle), *barâbar* (equal)

³¹ *câr* is a short form of *cahâr* (four)

³² This particular kind of construction is used in informal speech.

³³ *nam-nam e bârân* (drizzling rain).

Section 3.6.2 discusses the orthographic issues related to compound words in eFarsi.

3.6.2 Writing Compound Words

In some European languages, such as German and Swedish constituents of a compound are usually concatenated to form a single word, for example, the Swedish words *aktiebolag* [*aktie* (share, equity) + *bolag* (company)](limited company), *regeringspartiet* [*regering* (government) + *partiet* (the party)] (the ruling political party). In English, compound words could be written in three different formats: open format (spaced as separate words), hyphenated format (words separated by a hyphen) or a closed format (words concatenated to form a single word) [10]. The same sort of format is practiced in almost all Latin-based languages. Similarly, in eFarsi, we have chosen to classify the formatting of compound terms as follows - We select Persian names for the formats to avoid confusion.

- *Gosaste* (separate): In this case the words of a compound term are written separately using a space as a delimiter. For example, *Emârat e Arabi* (United Arab Emirates). Usually, the occurrence of the *Ezâfe* plays the role of a glue.
- *Nime-Peyvaste* (half-joined): In this format the words are also written separately, but a *xatt-e-fâsele* (hyphen) is used as a delimiter. For example, *xatt-e-fâsele*.
- *Peyvaste* (joined): In this format, the words of a compound term are concatenated to form a new word. For example, *golâb* (rose-water).

EOC 15 (*Compound Words*)

Compound words are written in *Peyvaste*, *Nime-Peyvaste* or *Gosaste*-format.

In the rest of this section a number of rules will be specified for those cases where the *Peyvaste*-format should apply. These rules however, may not be sufficient, in which case we propose the following general rule of thumb: as a word combination appears for the first time it is written according to the *Gosaste*-format and as it is used more often for a longer period of time, it moves to a *Nime-Peyvaste* format and finally to the *Peyvaste*-format. For example, the compound term *gol e yax* (sort of flower - 'ice flower') may after some persistent usage be written as *gol-e-yax* and as it is used more and more

one may write it as *goleyax*. Naturally, finalization of such formatting or indeed any of the proposed conventions of eFarsi is subject to the decisions made by authorities such as the Iranian *Farhangestân e Zabân va Adab e Fârsi* (in short *Farhangestân*).

A useful point to remember is that writing words in the *Peyvaste*-format may lead to some ambiguities or mispronunciations. For example, in the following cases:

- A prefix ends with the same letter as the initial letter of the prefixed word.
- A suffix starts with the same letter as the final letter of the suffixed word.
- The final letter of a prefix together with the initial letter of the prefixed word form a diphthong, for example, *beyâb* (find) that contains the *be*. The *ey* here is not an occurrence of the diphthong *ey*.
- The first letter of a suffix together with the final letter of the suffixed word form a diphthong. For example, suppose we use the diphthong *sh* instead of *sĥ*; the word *kushâ* could either be *kušâ* (Diligent) or *kus-hâ* (plural of *kus* (a music instrument)).

In some of these cases it may be more appropriate to write the compound word in a *Nime-Peyvaste* format. By so doing there will be no confusion or ambiguity.

In the following sections we will first list the possible word combinations and then explain the construction methods and finally provide some examples.

Simple concatenation

Simple concatenation of some nouns creates a compound word written in the *Peyvaste* format, for example, *bâqbân* [*bâq* (garden) -*bân* (-keeper)] (gardener), *serkeangabîn* [*serke* (vinegar) *angabîn* (honey)] (mixture of vinegar and honey - used as salad dressing)³⁴, *âbpâš* [*âb-pâš*] (watering-can)

³⁴*Serkeangabîn* has actually been transformed to *sekanjabîn* and is usually used in this form.

Using the connective *va* (or *o*)

As we saw in the previous section, a common way of constructing compound words is to join them with the connective *va* (and) or rather its short form *o*. We saw examples involving two verbs (2) and an infinitive and a verbal noun (8). Here are some examples (including some of the earlier ones):

jost o ju (search), *šost o šu* (washing), *goft o gu* (conversation), *kâr o bâr* (work), *âs o pâs* (hopeless), *rok o râst* (open and honest), *xord o xâb* (eating and sleeping), *pas o piš* (backward and forward), *kâr o bâr* (work), *âs o pâs* (hopeless), *seft o saxt* (solid and hard), *did o bâzdid* (visiting each other), *raft o âmad* (commuting), *kas o kâr*³⁵ (relatives and close people), *nâm o nešân* (name and address), *xert o pert* (junk, stuff), *cart o part* (irrelevant talk - bullshit), *kam o biš* (more or less)

Some of these compound terms have through time been transformed into a single word and even in the PA-Script are usually written as one word; furthermore, the connective *o* has been transformed into an *e* (This *e* should not be confused with *Ezâfe*). For example, *jost o ju* has been transformed to *josteju*, *šost o šu* to *šostešu* and *goft o gu* to *goftegu*. It is natural to continue writing these words in the *Peyvaste*-form in eFarsi. For the general case we think the connective *o* works fine as a glue and a *Gosaste*-format or a *Nime-Peyvaste*-format are appropriate, for example, *did o bâzdid* or *did-o-bâzdid*.

EOC 16 (*Compounds constructed with O*)

Compound words constructed by the preposition *o* (short form of *va* (and)) should be written in *Gosaste* or *Nime-Peyvaste*.

EOC 17 (*Compounds constructed with O - Exception*)

The following compound words may be exempted from EOC-16:
jostoju, *goftegu*, *šostešu*.

Compound words using *Ezâfe*

Ezâfe is a common means for constructing compound words in Persian. Just like *o* in the previous section, *Ezâfe* works fine as a gluing mechanism

³⁵Normally used in the negative form as *bi kas o kâr* meaning a person who doesn't have any relative/roots in a community.

in constructing compounds; therefore, writing these terms in the *Gosaste*-format is quite appropriate. For example, there is no reason to join the words of the compound word *Park e Mardom* (name of a park - people's park).

EOC 18 (*Compounds with Ezâfe*)

Compound words constructed by *Ezâfe* should be written in *Gosaste* or *Nime-Peyvaste* format with *Ezâfe* as a separate word.

However, dropping *Ezâfe* is a common phenomenon in Persian. Once *Ezâfe* is removed from a construction, the glue is lost and it is important to reconsider EOC-18 for such cases. In fact, a number of different cases arise.

1. ***Fakk e Ezâfe - Preserving the order of the Mozâf and Mozâfon Elayh.*** After dropping the *Ezâfe* from a construction, if the order of *Mozâf* and *Mozâfon Elayh* remains as before, we propose they should be written in the *Peyvaste*-form. For example, *pedarzan* [*pedar e zan*] (wife's father), *sarmâye* [*sar e mâye*] (capital), *pesarxâleh* [*pesar e xâle*] (son of mother's sister), *pedarbozorg* [*pedar e bozorg*] (big daddy, old daddy) (grandfather), *sâheb xâne* [*sâheb e xâne*] (owner of the house) (host). Words constructed this way, are suitable candidates as main entries in a lexicon.
2. ***Ezâfe ye Maqlub - Switching the order of the Mozâf and Mozâfon Elayh.*** After removing the *Ezâfe*, in some cases, the order of *Mozâf* and *Mozâfon Elayh* is reversed. The writing format depends on the meaning of the new construction. For example, *xâne ye mehmân* (house of guest) that is transformed into *mehmân xâne* (motel).
 - (a) The new sequence of the words *Mozâfon Elayh* and *Mozâf* does not lead to a linguistic unit with a meaning different from the original construction, both words keep their originally intended meaning and identity and remain in the same relationship as before. This is quite common in Persian poetry and also in some dialects of Persian it is normal way of assigning an adjective (*Mozâfon Elayh*) to a noun *Mozâf*.
In such cases, the words should be written separately (*Gosaste*-format).
 - (b) The word sequence resulting from *Mozâfon Elayh* followed by *Mozâf* results a new linguistic unit with a meaning different from

the original words (*Mozâf* followed by *Ezâfe* and *Mozâfon Elayh*). Words constructed in this way are natural candidates as main entries in a dictionary. In fact, this is a useful method for creation of new words.

In such cases, we think that the words *Mozâfon Elayh* and *Mozâf* should be concatenated to form a single word.

Some examples of 2a follows. Consider the following poem by the great Iranian poet and mathematician Xayyâm (Khayyam):

Pîs az man o to, leyl o nahâri budast
Gardande falak niz, be kêri budast

Har jâ ke qadam nahi to bar ru ye zamîn
Ân mardomak e cešm e negâri budast

gardande falak in the second line is a term that fits the case 2a described above. Normally, it would be written as *falak e gardande* (rotating heavenly wheel). But even though the order has changed and the *Ezâfe* is dropped, *gardande falak* has the same meaning as *falak e gardande*. Since this construction is only temporary, it will be inappropriate to write it as *gardandefalak*³⁶. Similar examples are: *sîb e sorx* (red apple) transformed into *sorx sîb*. Some compounds that have over time taken the *Peyvaste* format are, *seylâb* (flood water) *âb e seyl* (water of the flood), *xošhâl* (happy) *hâl e xoš* (happy mood), *âsiâsang*³⁷ (millstone) *sang e âsiâ* (millstone).

Examples for case 2b: *golbarg* (petal) *barg e gol* (leaf of flower), *šâhpesar* (distinguished or exceptional son or boy) *pesar e šâh* (son of the king), *mehmânxâne* (hotel, motel) *xâne ye mehmân* (guest's house), *gâvzabân* (borage) *zabân e gav* (cow's tongue), *narmafzâr* (software) *afzâr e narm* (soft tools), *xošraftâr* (well-behaved person) *raftâr e xoš* (pleasant behaviour), *ostâdyâr* (assistant professor) *yâr e ostâd* (friend/lover/assistant of the master).

Adding Prefixes

In eFarsi, all prefixes are simply concatenated to the beginning of the following word in the *Peyvaste* format. There are a large number of prefixes in

³⁶And most probably Xayyâm didn't mean to create a new word here.

³⁷*Âsiâsang e zirîn motaharrek nîst, lâjeram tahammol e bâr e gerân konad.* Sa'di

Persian. Some common prefixes are: *mî*, *be*, *na*, *bar*, *bî*, *nâ*. Some compound words containing these prefixes are: *mîguyad* (is saying), *begu!* (say!), *nagu!*, *don't say*, *bargašt* (return, returned), *bigonâh* (innocent), *nâdân* (stupid).

EOC 19 (*Prefixes*)

A prefix and the prefixed word are written in the *Peyvaste*-form.

Adding Suffixes

All suffixes are simply concatenated to the end of a word and we suggest that the suffix and the word should be written in the *Peyvaste* format unless there is a risk of mispronunciation. In Farsi, there are a large number of suffixes: *-hâ* or *-ân* for constructing plurals, *-i* for transforming an indeterminate noun to a determinate noun, *-mand* for assigning ownership. For example, *setârehâ* (plural of *setare* (star)), *abruvân* (plural of *abru* (eyebrow)), *pâyhâ* (plural of *pâ* (foot, leg)), *mardi mard-i* (a man), *danešmand daneš-mand* (scientist), *xeradmand xerad-mand* (wise).

EOC 20 (*Suffixes*)

A suffix and the suffixed word are written in the *Peyvaste*-form.

Adding an infix

Infixes are not so common in Persian, but they exist. The appropriate format for compound words with an infix is the *Peyvaste*-format. For example, *â* in *kešâkeš* (struggle), *peyâpey* (one after the other, in series), *banâguš* (cavity behind the ear), *takâpu* (search, running about), *zanâšui* (Matrimony).

EOC 21 (*Infixes*)

An infix and the words surrounding it are written in the *Peyvaste*-form.

3.6.3 Exceptions

Ezâfe

As it was mentioned earlier, *Ezâfe* is always written separately except in cases such as *barâye* (for) or *bedune* (without). Such words almost always appear together with the *Ezâfe* and it is for this reason we recommend joining the *Ezâfe* to the word.

EOC 22 (*Ezâfe Exceptions (see EOC-11)*)

If Ezâfe for some reason can not be separated from a word then it is joined to it. Typical examples are: *barâye* (for), *bedune* (without).

In and *Ân*

Vâzehâ ye esâre (demonstratives), *in* (this) and *ân* (that) are always written separately, except in the following cases where they are glued to a neighbouring word: *injâ* (here), *ânjâ* (there), *ânçe* (that which), *inhâ* (these), *ânhâ* (those), *inke* (this who), *ânke* (that who), *hamin* (this), *hamân* (that very), *conin* (such, such a one), *conân*³⁸ (such, such a one), *inconin* (such a one), *ânconân* (such a one)³⁹, *hamconin* (just like this, also), *hamconân* (like before, as it has been, still), *candin* (several, so many), *candân* (so many, so much)⁴⁰, *intour* (thus, in this way), *ântour* (in that way), *ingune* (in this way), *ângune* (in that way), *ângâh* (then).

Be

The preposition *be* (to) is written separately, for example, *Mâ be sinemâ raftim* (We went to the movies), *xâne be xâne* (house to house), *be âTehrân raft* (sHe went to Tehran). In the following cases, however, *be* is should be written in the *Peyvaste*-format:

- In *bejoz* (except)
- When *be* is written as *bed* in, and only in, the following words: *bedin*, *bedân*, *bedu*, *bedîshân*.
- *be* is sometimes used as the so called *Be ye Zinat* (the redundant *be*) at the beginning of a verb or an infinitive. For example, *begoftam* (same as *goftam* (I said)), *begoftan* (to say). If the verb following *be* starts with an *a* or an *â*, then the *be* is transformed to *bi*, for example, in *biafkanad* [*be+afkanad*] (throws), *biâ* [*be+â*] (come).
- *be* can be added to a noun to create an adjective or an adverb, for example in *benâm* (famous) where it is added to *nâm* (name); and in *besor'at* (fast) where it prefixes *sor'at* (speed)

³⁸ *conin* is also pronounced as *cenin* and *conân* is written as *cenân*. The *o* in *con* is pronounced as *e* in other similar constructions.

³⁹ *Ânconân oftâd ke pâyaş şekast*.

⁴⁰ *Candân ham gerân nabud* (It wasn't so expensive).

Bâ

Bâ (with) is written separately unless it is used to create a compound adjective by prefixing a noun, for example, *Mâ bâ u dustim* (we are her/his friends), *bâhonar* (talented), *bâhuş* (clever).

Bâz

Bâz (again) is written separately unless used as a prefix, for example, *Bâz be mâ sar bezan* (Do come and visit us again), *bâzgaştan* (return), *bâzdâst* (arrest, detention), *bâzpors* (interrogator).

Bi

Bi (without) is written separately unless it is used to create a negative adjective, for example, *Man bi to be kojâ beravam* (Where shall I go without you; I'll be lost without you), *bihameciz* (), *bişilepile* ()

Ce

Ce (what, thing) is written separately, except in the following cases: *ânce* (that which'), *conânce* or *cenânce* (if indeed), *cerâ* (why), *cehâ* (plural of *ce* (what)), *cetowr* (how), *cetowr* (how, in what way), *cegune* (how, in what way)

Ke

Ke (who, that) is written separately, except in the following cases: *inke* ('this who', 'this which'), *ânke* (that who), *kehâ* (plural of *ke*), *conânke cenânke* or (if indeed), *haminke* (as soon as)

Râ

Râ is a preposition that follows a definite noun and marks it as the direct object of the verb, providing a pointer to person or thing the verb is being applied to. It is written separately, except in the following cases: *cerâ* (why)⁴¹ *marâ* (short for *man râ* (me)), *zirâ* (because)

⁴¹*cerâ* should not be confused with *ce râ*. Note the difference between *cerâ xând?* (why did he read) and *ce râ xând?* (what did he read?).

Ham

Ham appears both as a prefix and as a word. As prefix, it follows the general rule for prefixes and is written in the *Peyvaste*-format. In all other cases, *ham* is written separately. For example, *Hammihan* (fellow countryman), *hamkelâs* (classmate), *hamkâr* (co-worker), *hamdigar* (each other), *hamrâh* (fellow traveller, escort), *hamcenân* (as usual), *hamin* (this one), *hamân* (that one), *hamânâ* (those ones), *bâ ham âmadim* (we came together), *to ham biâ* (you come too).

3.7 Geographic Names, Proper Names and Trademarks

The transliteration conventions of eFarsi do not apply to trademarks. In general, names of people and places are also exempted from these conventions. For example, trademarks such as SMIRNOFF, LINUX, VOLVO, etc. must be written as they are and not according to how these words are pronounced in Persian. The same is true for names as well, for example, a reasonable way of writing 'Geoffrey' in eFarsi would be as *Jefri*, but somehow it is wrong to change the spelling of the name. In fact the most appropriate way of writing peoples names is to write them according to the wishes of the owner of the name. For example, we would write 'k.d. lang' and not 'K.D. Lang' in eFarsi.

Another issue related to names is that, in spoken Persian, an Ezâfe is placed between the first name and the surname of a person unless the first name ends with a *â*. It is quite reasonable to exclude this Ezâfe in writing, so we will write *Shirin Ebâdi* although her name is pronounced as *Shirin e Ebâdi*.

Finally, in case of names (personal or geographic) we would like to distinguish between those words that have a well-established Persian pronunciation and others. For example, *Landan* is the established Persian name for London and our intention is to accept *Landan* as the eFarsi word for London. In the next section (Section 3.8), we will sketch a general rule for dealing with non-Persian words.

3.8 Foreign words

In a proposal such as eFarsi, dealing with foreign (non-Persian) words is a complicated issue. At the same time that we strive for general writing rules, we need to respect well-established usage of words. Given a foreign word *W* that belong to a language *L* with a script *S*, we sketch the following rule for writing the eFarsi version of *W*:

- If *W* has a well-established pronunciation in Persian then the eFarsi version of the word would be based on that pronunciation, for example, we will write *Pâris* instead of *Paris*, *Kubâ* instead of *Cuba*; otherwise,
- if the *S* is based on the Latin alphabet, then the spelling of the original language *L* would apply, for example, the English word 'Sir' and the name of the Swedish city 'Grnna' would be written in exactly the same way in eFarsi; otherwise,
- if *S* is not based on Latin but there is a commonly used (or standard) Latin-transliteration for it, then the spelling suggested by the Latin-transliteration would be the eFarsi-spelling for *W*; otherwise,
- a suitable eFarsi spelling, based on the pronunciation of the word in the original language, should be produced, for example, *Na'am* (Yes in Arabic).

3.9 Some Modifications and Extensions

3.10 Making eFarsi More Practical

In this section we introduce some diphthongs that can be used as alternatives to some of the letters of the eFarsi alphabet. In general, having more than one alternative letter for a sound is not a wise choice, but as long as there is no simple way of entering letters such as \hat{a} , \hat{s} and \hat{z} (or their equivalents in similar proposals [3, 4]) into your computer, having a standard alternative is appreciated. It is mainly therefore we suggest the diphthongs *aa*, *sh* and *zh*. We also introduce three more diphthongs which are quite common in traditional transliterations: *ch*, *gh* and *kh*.

EOC 23 (*Diphthongs Are Practical*)

aa, *ch*, *gh*, *kh*, *sh* and *zh* can be used as alternatives to \hat{a} , *c*, *q*, *x*, \hat{s} and \hat{z} respectively.

aa is an alternativ for \hat{a} , for example, *baayad* (must), *aayaa* (whether).⁴²

ch is an alternativ for *c*, for example, *che* (what), *gach* (chalk).

gh is an alternativ for *q*, for example, *gham* (sorrow).

kh is an alternative for *x*, for example, *Khayyam* (Great Persian poet and mathematician).

sh is an alternativ for \hat{s} , for example, *shab* (night), *shode* (has become).

zh is an alternativ for \hat{z} , for example, *Bizhan* (Persian name for men, also written as Bijan).

EOC 24 (*The Extended eFarsi Alphabet*)

Here is a complete list of the eFarsi letters including the diphthongs: \hat{a} , *aa*, *a*, *b*, *c*, *d*, *e*, *ey*, *f*, *g*, *gh*, *h*, *i*, *j*, *k*, *l*, *m*, *n*, *o*, *ow*, *p*, *q*, *r*, *s*, *sh*, *t*, *u*, *v*, *w*, *x*, *kh*, *y*, *z*, *zh*, ' ,

3.11 Improving the look of the words

Having a simple transliteration scheme is great. But making the words nicer is also important. In this section we introduce some exceptions to the general rules in order to improve the look of certain words.

⁴²Yet another simplification would be to use *a* for \hat{a} . That would mean that *bad* (bad) and *bâd* (wind) would both be written as *bad* but often the context helps the reader to distinguish the two. In a number of European languages, for example, English and Swedish both long and short a-sound is written as 'a'.

3.11.1 *Taâdid* on Certain Diphthongs

We mentioned earlier that due to limitations of computer keyboards in different countries or due to common practice, some diphthongs could replace certain letters used in eFarsi. For example, *sh* was mentioned as an alternative to *sh*. This would mean that the word *baâshâsh* (bright) would have to be written as *bashshaash*. It is very tempting to remove the first occurrence of the letter *h* and by so doing improving esthetical aspects of the word. Compare also *bachche* (child) with *bacche*, *oshshâq* (lovers) with *osshâq*, and *saghghez* (chewing gum) with *sagghez* which are alternatives to *bacce*, *osshâq* and *saqqez*. Since we are not in the process of creating a standard, we could only suggest this as a proposal. The only tangible evidence supporting this *h*-dropping operation is that two individuals interested in Persian transliteration and transcription have suggested it independently. We considered the issue in 2002 but later we learned that [11] had suggested removal of the first *h* as a possible orthographic decision many years ago.

3.11.2 *Taâdid* on *Ist*

Ist, the glottal stop in eFarsi is represented by a single quote and is used in the transliteration of the Arabic letters **Eyn** and **Hamze**. In some words, there is a ***Taâdid*** on these letters, for example, the Arabic word *fa'âl* (active). In eFarsi transliteration we have decided to reduce the supposedly double occurrence of *Ist* to only one since there seems to be no loss of information, *fa'âl* thereby is written as *fa'âl*.

3.11.3 *Ist* Between Vowels

In Persian, the occurrence of *Ist* between two vowels is for indicating a change in vowels, for example, *tavânâ'i* (capability), *pâ'iz* (autumn). Explicit mention of *Ist* in the PA-Script is important since vowels are usually left out in writing. In eFarsi, however, vowels are explicit, and therefore making the *Ist* that occurs between two vowels redundant. For example, we could write *tavânâi* instead of *tavânâ'i*. Here are some examples:

jâme'e (society), *sâ'at* (watch, clock, hour), *so'âl* (question),
masâ'el (problems), *fa'âl* (active)

Compare these with:

jâmee (society), *sâat* (watch, clock, hour), *soâl* (question), *masâel*
(problems), *faâl* (active)

Essentially, the *Ist* in these cases ensures that a vowel preserves its sound even if it is preceded or followed by another vowel. Consider, for example, the English words 'koala' and 'coal'. The occurrences of 'oa' in these two words have different pronunciations, the first 'a' in 'koala' is more distinct, whereas the 'a' in 'coal' is not pronounced and is simply used to modify the sound of the 'o' that precedes it.

In eFarsi, the sound of each vowel appearing in a word is almost completely independent of other letters in the word. A vowel is not used to modify the sound of another vowel.⁴³ In some Persian dialects dominant in Iran, the *Ist* is often replaced with a prolonged pronunciation of some neighboring vowel in order to compensate for the *Ist*. This vowel, however, is not always an immediate neighbor of the *Ist*. Consider the following examples in Tehrâni Persian, where the prolonged vowel is overlined:

- The *Ist* in *sham'* (candle) is replaced by prolongation of the vowel *a*, like so: *shām*
- The *Ist* in *ro'b* (fear) is replaced by the prolongation of *o*, like so: *rōb*

This vowel prolongation phenomenon appears to be a much simpler task for the speech apparatus of (Tehrâni) Persian speakers than *Ist* is.

3.12 Capitalization, Abbreviation and Punctuation

3.12.1 Capitalization

Since capitalization of certain letters improves the readability of text, in line with many other Latin-based scripts, we propose similar rules for it. Here is a list of cases where capitalization is useful.

1. Capitalization should respect the rule for writing names of people and trademarks which is specified in Section 3.7. As stated, trademarks and other registered names should appear exactly as the specifications of the trademark dictates. For example, CNN, COCA COLA, AOL, UNESCO, cK
2. The first word of a sentence is capitalized. For example, *Yek sâ'at digar otobus miresad* (The bus arrives in an hour).

⁴³This is not to say that the pronunciation of two neighboring vowels do not affect each other at all. For example in section 4 where some common mistakes are discussed, we see that the pronunciation of *iâ* in some words has lead some to mistakenly use *iyâ* for achieving the sound effect that results when occurrence of *iâ* in some words is pronounced.

3. The first word of a syntactically complete quoted sentence is capitalized. For example, *Mahnâz goft, 'Barâbari e zan va mard dar jâme'e barâye man mohemm ast'* (Mahnâz said, 'Equality between men and women is important for me')
4. Proper names and geographic names are always capitalized. If these names are compound names then every 'major' component should be capitalized. For example, *Tehrân* (Tehran - capital of Iran), *Iâlât e Mottahed e Âmrika* (United States of America), *Sâsmân e Melal e Mottahed* (United Nations), *USA* (USA), *Ra'is e Koll e Qovâ* (Head of the military forces - ground force, air force and the marines), *Majles e Sôwrâ ye Eslâmi* (the name of the current Iranian parlement). Even certain expressions that refer to well-defined geographic parts of the world or phenomenon should be capitalized, for example, *Xâvar e Dur* (Far East), *Bâd e Sômâl* (Northern Wind). But compass directions in general are not capitalized, for example, *jonub e qarbi e Irân* (south west of Iran).
5. Contrary to the practice in English, the names of weekdays, months, and years are not capitalized. For example, *došanbe* (Monday), *farvardin* (First month of the Iranian calendar), *ut* (August), *sal e meymun* (year of the monkey)
6. Abbreviated titles are always capitalized. For example, *Âq. Rezâ Âsuri* (Mr. Rezâ Âsuri), *Dr. Ahmadi* (Dr. Ahmadi), *Xâ. Simâ Sîrâzi* (Mrs. Simâ Sîrâzi), *Du. Lâle Kermâni* (Miss. Lâle Kermâni)
7. In articles and books, all main words appearing in a title or chapter names are capitalized, for example, *Fasl e Yek: Zabân e Fârsi* (Chapter One: Persian Language)
8. Nationality should be capitalized, for example, *dânešju ye Irâni* (Iranian student), *hame ye Irâniân* (all Iranians). This should also apply to other geographic units, for example, *bâzârhâ ye Âsiâi* (Asian markets)

3.12.2 Abbreviation

We consider the following types of abbreviations.

1. Abbreviation of single words should follow an standardization. These abbreviations should end with a period (.). For example, *Teh.* as a

possible abbreviation for *Tehran, Dkade. Barq* (Dept. of Electrical Engineering) as a possible abbreviation for *Dâneškade ye Barq, q.* (adv.) as a possible abbreviation of *qeyd* (adverb)

2. When abbreviating a compound name, the first letter of each major word in the compound name is included in the abbreviation, for example, *SNI* as the abbreviation of *Ŝerkat e Naft e Irân* (Iranian Petroleum Company), *RI* as the abbreviation of *Râdio Irân* (Radio Iran)
3. When abbreviating an expression or a construction consisting of two words or more, the first letter of each major word is included in the abbreviation and succeeded by a period. For example, *b.m.* as an abbreviation for *barâye mesâl* (for example), *v.e.a. va elâ âzar* (and so on, etc.), *b.b.i.* as an abbreviation of *banâ bar in* (therefore, hence).
4. Titles should normally be abbreviated, for example, *Âyatollâh Taleqani* (Ayatollah Taleqani) could be written as *Âyat. Tâleqâni*, *Dușize Leylâ Golcin* (Miss. Leylâ Golcin) could be abbreviated as *Du. Leylâ Golcin*, *Porofesor Pari Mehrabân* (Professor Pari Mehrabân) could be abbreviated as *Porof. Pari Mehrabân*
5. It is quite practical to have abbreviations for ordinal numbers, like the English first, second, third, fourth, and so on, which in a linear flow of text are written as 1st, 2nd, 3rd, 4th, ... and as 1st, 2nd, 3rd, 4th, ... in other texts such as mathematical texts.

In Persian, ordinals (or ordinal numbers) are constructed by adding the suffix *-om*⁴⁴ or *-vvom* to a number. For example, *yekom* (first), *dovvom* (second), *sevvom* (third), *cahârom* (forth), and so on.⁴⁵

We propose that these numbers be abbreviated by writing the numeral followed by the suffix *-om*, like so: 1om (first), 2om (second), 3om (third), 4om (forth), ..., or as 1^{om}, 2^{om}, 3^{om}, 4^{om}, ..., in mathematical text.

Another sequence of numbers for ranking in Persian is: *yekomin*, *dovvomin*, *sevvomin*, *cahâromin*, ... The difference between, *sevvom* and *sevvomin*, for example, is that *sevvom* appears after the word whose order is being given and *sevvomin* occurs before the word (see

⁴⁴When the number is not known, *cand* is used instead and the suffix *-om* is added to it: *candom*.

⁴⁵As you may have noticed, the suffix *-vvom* is used for numbers that end with a vowel and *-om* is used when the name of the number ends with a consonant.

the discussion on ***Ezâfe ye Maqlub*** in Section 3.6). The following expressions are equivalent:

Konferâns e Sevvom e Anjoman e Ânformâtik e Irân
Sevvomin Konferâns e Anjoman e Ânformâtik e Irân

As noted, in the second case the order of the rank and the word being ranked is exchanged and the *ezâfe* is dropped.

For these numbers we suggest the following abbreviations: 1in, 2in, 3in, 4in, . . . , or 1ⁱⁿ, 2ⁱⁿ, 3ⁱⁿ, 4ⁱⁿ,

3ⁱⁿ Konferâns e Anjoman e Ânformâtik e Irân
 (3rd Conference of the Informatics Society of Iran).

6. – (that is, *xatt e fâsele* (dash sign)) can be used as an abbreviation of the word *tâ* (to) which is used to specify intervals, for example, *s. 11-23* as an abbreviation of *az safhe ye 11 tâ 23* (from page 11 to 23), *došanbe–panjšanbe* (Monday–Thursday). In the same way, we abbreviate the connective *va* (and), for example, *didâr e Khatami–Adnan* (Khatami-Adnan meeting), *ravâbet e Irân–Âmrikâ* (Iranian-American relations).

3.12.3 Punctuation

eFarsi and PA-Script, in principle, follow the punctuation rules that are practiced in most European languages. We mention some of these conventions, but for a better coverage of the topic the reader should refer to standard texts for editors and writers (e.g. [10]).

1. A statement (*jomle*) ends with a period (.).
Man târix e Irân râ balad nistam. (I do not know Iran’s history.)
2. A question (*porseš*) ends with a question mark (?). *Esmet chie?* (What is your name)
3. An exclamation (*jomle ye ta’ajjobi*) ends with an exclamation mark (!). *Mage miše târix e kešvarat râ balad nabâsi* (How could you not know your country’s history!)
4. An imperative (*jomle ye amri*) ends with an exclamation mark (!).
Parisâ, lotfan otâqat râ tamiz kon! (Parisâ, please clean your room!)

5. In order to specify a list of items in the text flow, end the word preceding the list with a colon (:) and separate the items with a comma. For example,

Yek hafte haft ruz dêrad: ânbe, yekânbe, doânbe, seânbe, cahârânbe, panjânbe, âdine (or jom'e) (A week has seven days: Saturday, Sunday, Monday, Tuesday, Wednesday, Thursday, Friday)
6. Use commas to improve readability of the text. For example, *Barâye mesâl, farvardin 31 ruz ast va mehr 30 ruz* (For example, month of farvardin is 31 days and mehr 30 days)
7. Quoting rules are as usual, for example, *Âraê goft: 'Midâni "omni-science" be Fârsi ci miêe?'* (Âraê said: 'Do you know what "omni-science" would be in Persian?')

4 Common Mistakes

In this section we briefly discuss some of the common transliteration mistakes. In particular, the letter *i* followed by other vowels *â*, *a*, *e*, *i*, *o* and *u* seems to be a major problem. Usually, *y*-sound between the pronunciation of *i* the other vowel. The fact is that the brief *y*-sound is nothing more than a side effect of trying to pronounce *i* and the other vowel by softly moving from the sound of one to the other. Some English speakers would, for example, pronounce 'vodka and lime' as 'vodkar and lime'. The appearance of 'r' in this context is very similar to the appearance of the *y*-sound. For example, when pronouncing *siâh* it seems that we are pronouncing the word *siyâh*.

Some more examples are provided in table ??.

Correct	English	Incorrect
<i>xiâl</i>	imagination	<i>xiyâl</i>
<i>piâde</i>	(on foot, pedestrian)	<i>piyâde</i>
<i>niâkân</i>	ancestors	<i>niyâkân</i>
<i>siâh</i>	black	<i>siyâh</i>
<i>farhangiân</i>	academics	<i>farhangiyân</i>
<i>vidio</i> or <i>video</i>	video	<i>vidiyo</i>
<i>televizion</i>	television	<i>televizyon, televiziyon</i>
<i>râdio</i>	radio	<i>râdiyo, râdyo</i>
<i>kâmpiuter</i>	computer	<i>kâmpiyuter, kampyuter</i>
<i>kiân</i>	a male name	<i>kiyân, keyân</i>
<i>pâiz</i> or <i>pâ'iz</i>	autumn	<i>pâyiz, pâeiz</i>
<i>sui</i> or <i>su'i</i>	a direction	<i>suyi, suey, suei</i>
<i>rais</i> or <i>ra'is</i>	boss	<i>rayis, raeis</i>
<i>said</i> or <i>sa'id</i>	a male name	<i>saeed, saeid</i>

5 Discussion and Further Work

There are no standard orthographic principles for writing Persian using the Latin script. There are a number of proposals but none of has given any rigorous account of transliteration of Persian to the latin alphabet. We hope that the orthographic principles we have sketched here eventually contribute towards a standard. But actual standardization is a complicated process beyond the scope of this article.

eFarsi is an orthographic scheme. As well as introducing an alphabet, we have introduced a number of conventions and, where necessary, provided the reasons for the choices that have been made. A lot of work remains to be done. There is a need for computer readable dictionaries and resources. Furthermore, we need software for automatic conversion between Persian text written in the PA-Script and eFarsi.

We think that eFarsi is a natural complement to the existing writing systems in the Persian-speaking countries. Even without enforcing a Latin alphabet, there are a number of different uses for a Latin alphabet.

5.1 Why Latin?

Our of the reasons for proposing eFarsi is economical. It is so much easier to keep up with technological development in text-processing once we use the mainstream alphabet of the world, Latin. Since eFarsi preserves the

'usual' pronunciation of most of the Latin letters, it will be easier to learn the alphabet and use it effectively.

5.2 Further Issues

There are some further issues that we are aware of, but we have decided to postpone treatment of these issues. Here are some of them:

- Just like English, Persian has two I-sounds: long and short. English mostly employs the diphthongs *ee* and *ea* for the long i-sound, and the letter *i* and sometimes the letter *y* for the short i-sound⁴⁶, for example, *been*, *seen*, *deed*, *teacher*, *bit*, *rythm*, *clergy*.

For the sake of simplicity, we have proposed one i-sound for both the long and the short i-sounds, both transliterated as *i*. As we have mentioned earlier, the short i-sound does not occur so often in Persian. However, further discussion on this issue is justified.

- We have decided to write the *Ezâfe* separately, for example, *nur e xorâsid* (light of the Sun), *pâ ye to* (your foot). This is also suggested in [3], but [7] and [4] propose concatenation of the *Ezâfe* to the end of the first word (***Mozâf***), for example, *nure xorâsid*, *pâyef to*. We think the choice we have made in eFarsi is wise but we hope future discussions and actual usage of these conventions would shed more light on the advantages and disadvantages of each alternative.
- The PA-letters *Qeyn* and *Qâf* have almost the same sounds in Persian. In Arabic, however, there is a clear difference in the pronunciation of the two. Both *Qeyn* and *Qâf* are transliterated to *q*, but some may argue that they should also be distinguished in schemes such as eFarsi. There could be various arguments for so doing. Some may argue that *Qeyn* is a genuine Persian letter and therefore it deserves its own orthographical representation (for example, *ġ*). Some others may argue that since *Qâf* only appears in words that are originally Arabic and furthermore, Persian-speakers pronounce it a bit differently anyway, then it is important to have two different orthographical representations for the two.

But there seems to be some sort of consensus for the time being that a single letter (*q*) is sufficient for transliteration of *Qeyn* and *Qâf* [3, 4, 7].

⁴⁶Actually this is too much of a simplification since things are much more complicated in English.

Appendix A

This appendix provides the exact pronunciation of the eFarsi letters using the international phonetic system. To be completed !!!

Appendix B

There are various sorts of *Ezâfe* in Persian. For a complete list the reader should refer to a Persian grammar book.

1. *Ezâfe ye Melki* (possessive *Ezâfe*) where *Mozâfon Elayh* owns the *Mozâf*, for example, *xâne ye mâ* (our house). In this case *Mozâfon Elayh* is always a person (capable of possessing).
2. *Ezâfe ye Taxsisi* (Allocational *Ezâfe*) where the *Ezâfe* allocates *Mozâf* to the *Mozâfon Elayh*, for example, *dar e otâq* (door of room), *rang e sîse* (colour of glass), *mardom e Soed* (people of Sweden). *Ezâfe ye Taxsisi* is very similar to *Ezâfe ye Melki*. The only difference is that in the latter *Mozâfon Elayh* is a person.
3. *Ezâfe ye Bayâni* (Descriptive *Ezâfe*) where the *Mozâfon Elayh* describes the type of *Mozâf* or the material the *Mozâf* is made of, for example, *kâse ye mes* (silver bowl), *deraxt e âlu* (plum tree), *samâvar e barqi* (electric samovar)
4. *Ezâfe ye Towzihi* (Explanatory *Ezâfe*) where the *Mozâf* declares the type of *Mozâfon Elayh* or in other words *Mozâfon Elayh* is an instance of the general concept represented by *Mozâf*, for example, *Rud e Karun* (Karun River), *Ŝahr e Tehran* (Tehran City), *ruz e ŝanbe* (day of Monday).
5. *Ezâfe ye Tašbihi* (Analogical *Ezâfe*) where analogy between the *Mozâf* and *Mozâfon Elayh* is drawn, for example, *mahd e zamîn* (mother Earth), *qad e sarv* (tall as a cypress), *lab e la' l* (ruby lips).
6. *Ezâfe ye Este'âri* (Metaphorical *Ezâfe*) where *Mozâf* takes a metaphorical role, for example, *dast e ruzgâr* (hand of time), *guš e došman* (enemy's ear), *Dast e sabâ be zolf e saman ŝâne mizanad - Abr e bahâr, na're ye mastâne mizanad*.
7. *Ezâfe ye Farzandi* (Sonship/Daughtership *Ezâfe*) where *Mozâf* is the son or the daughter of *Mozâfon Elayh*, for example, *Rostam e Zâl* (Rostam son of Zâl), *Mohammad e Abdollâh* (Mohammad son of Abdollâh - in Arabic *Mohammad ebne Abdollâh*), *Isâ ye Maryam* (Jesus son of Maria).
8. *Ezâfe ye Rabti* (Prepositional *Ezâfe*) where *Mozâf* is a primitive adverb or noun. For example, *bedune to* (without you), *birun e dar* (outside the door), *zir e miz* (under the table)

Appendix C: More Examples

This appendix would be extended.

â: *âb* (water) *bârân* (rain) *bâdâm* (almond)
a: *abâ* (fear) *tars* (fear) *vahsât* (fear) *na* (no).
b: *bâmâd* (morning) *abr* (abr) *âb* (night)
p: *parastu* () *sepâs* (thanks) *tup* (ball)
t: *tešne* (thirsty) *âtaš* (fire) *sokut* (silence)
s: *Sorayyâ* (Soraya) *hasti* (existence) *pârs* (Persia)
j: *jahân* (world) *vojud* (existence) *mowj* (wave)
c: *cehre* (face) *kuce* (alley) *kuc* (migration)
ch: *chehre* (face) *kuche* (alley) *kuch* (migration)
h: *huš* (intelligence) *nahân* (hidden) *panâh* ()
x: *xâk* (soil) *saxt* (difficult) *kâx* (palace)
d: *dâneš* (science) *pedar* (father) *šâd* (happy)
z: *zendegi* (life) *ruzgâr* () *râz* (secret)
r: *ru* (on) *šegarf* () *bahâr* (spring) *zâr*: *zârf* (depth) *pazuheš* (research) *kuž*

()

zh: *zharf* (depth) *pazuheš* (research) *kuzh* (
š: *šâdi* (happiness) *xošnud* (happy, satisfied) *soruš*
sh: *šâdi* (happiness) *xošnud* (happy, satisfied) *soruš* ()
q: *qâli* (carpet) *meqdâr* (amount) *barq* (electricity)
gh: *ghâli* (carpet) *meghdâr* (amount) *bargh* (electricity)
f: *farhixte* (knowledgable) *afsâne* (tale) *gazâf* ().
k: *kušâ* () *niki* (goodness) *pâk* (clean).
g: *gol* (flower) *hengâm* (at the time) *barg* (leaf).
l: *lagan* (pott) *pahlavân* (champion) *bâl* (wing)
m: *mast* (drunk) *peymân* (treaty) *ârâm* (still, quiet).
n: *nasim* (breeze) *minu* (sky) *javân* (young).
v: *vojud* (existence) *mive* (fruit) *sarv* (cypress tree).
y: *yazdân* (God) *peyvand* (connection, relation) *ra'y* (vote, will).
ow: *owlâd* (offsprings - arabic word) *Mowlânâ* (alias for Rumi) *now*
(new).

And finally, a poem by the great Iranian poet Xayyâm:

tâ chand hadis e panj o câr ey sâqi
moškel che yeki che sad hezâr ey sâqi

xâkim hame, chang besâz ey sâqi
bâdim hame, bâde biâr ey sâqi

References

- [1] Axondzâde, Mirzâ Fath-Ali, Biography, 1872.
- [2] Dedâ yat, Sâ deq, ???, 1945.
- [3] <http://www.eurofarsi.com>.
- [4] <http://www.iranianlanguages.com>.
- [5] Kasravi, Ahmad, ???.
- [6] Iranian Academy of Persian Language and Literature, *Dastur-e Khatt-e Farsi (Persian Orthography)*, Supplement No. 7, Nâme-ye Farhangestân, Tehran, Feb. 2000.
- [7] Persian Grammer, Cambridge University Press, 1953.
- [8] Maleki, J., Salehi, I., eFarsi: A Transliteration Scheme for Farsi, Working Paper, Dept. of Computer and Information Science, Linköping University, Sweden, 2001.
- [9] O'Grady William, Dobrovolsky Michael, and Aronoff Mark, *Contemporary Linguistic - An Introduction*, St. Martin's Press, New York, 1993.
- [10] The Oxford Guide to Style, Ritter, R. M., Oxford University Press, 2002.
- [11] Sanamrad, M. R., Open Letter, 1997.
- [12] Shari'at, Mohammad Javâd, Dasturabân e Fârsi, 1375 Hegri e Shamsi.
- [13] Transliteration of Farsi names for people and geographical locations in English subtitles of programs, IRNA: Official Iranian Radio and Television, 2002.
- [14] Yâsami, Rašid; Bahâr, Malekoššo'arâ; Foruzânfar, Badi'ozzamân; Qarib, Abdol'azim; Homâyi, Jalaladdin: *Dastur e Zabân e Panj Ostâd*. Ramin. 3rd Edition. Tehran 1378.