

eFarsi - A Latin-Based Writing Scheme for Farsi

Jalal Maleki

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

Email: `jma@ida.liu.se`

Abstract

This paper proposes a writing scheme for transliteration of Farsi 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*

Contents

1	Introduction	2
1.1	Farsi-Arabic Script	3
1.2	General Characteristics of eFarsi	5
2	eFarsi Alphabet	6
3	eFarsi Conventions	17
3.1	<i>Ist</i>	17
3.2	Silent <i>h</i>	18
3.3	<i>Taâdid</i>	18
3.3.1	<i>Taâdid</i> on the Farsi-Arabic letter <i>Ye</i>	19
3.3.2	<i>Taâdid</i> on the final letter	19
3.4	Ezâfe	19
3.5	Compound Words	22
3.5.1	What Words Are Combined	23
3.5.2	Building compound words	24
3.6	Exceptions	28

3.7	Geographic Names, Proper Names and Trademarks	30
3.8	Foreign words	30
3.9	Some Modifications	31
3.10	Making eFarsi More Practical	31
3.11	Improving the look of the words	31
	3.11.1 <i>Taâdid</i> on Certain Diphthongs	32
	3.11.2 <i>Taâdid</i> on <i>Ist</i>	32
	3.11.3 <i>Ist</i> Between Vowels	32
3.12	Capitalization, Abbreviation and Punctuation	33
	3.12.1 Capitalization	33
	3.12.2 Abbreviation	34
	3.12.3 Punctuation	36
4	Common Mistakes	37
5	Discussion and Further Work	37
	5.1 Why Latin?	38
	5.2 Further Issues	38
6	Appendix	39

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, Farsi is written in a variation of the Arabic script that we will refer to as the *Farsi-Arabic Script* (FA-script hereon) which is mainly used in Iran. Similar scripts are used in Afghanistan, Pakistan and parts of India. Farsi 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 teranslitereyâsen âv Fârsi tu Latin. Dis eskim, kâld eFarsi, kânsists âv an alfâbet and e nâMBER âv kânvensens dat perovâyd 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 Farsi but do not have good command of the FA-script. There are also millions of Farsi speakers who do not have access to Farsi-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 Farsi 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 Farsi script [3]. In [5], Lambton briefly describes a transliteration scheme for Persian and uses it throughout his book. There are also a number of recent proposals such as Eurofarsi [1] and Unipers [2] 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 Farsi-speaking countries.

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

Finally, although this paper assumes some knowledge of Farsi and the FA-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 alphabet is introduced and then some special cases that require more explanation and treatment are discussed. The appendices contain miscellaneous information. Throughout this paper the eFarsi text is printed in italics and Farsi or Arabic words are printed in bold italics.

1.1 Farsi-Arabic Script

For many centuries a variation of the Arabic alphabet has been used by Iranians to write Farsi. This alphabet includes all Arabic letters together

with four Farsi-specific letters.

Unfortunately, we do not have access to the FA-letters, 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 Farsi. And also the nine letters **Se, He-ye-Hotti, Zâl, Sâd, Zâd, Tâ, Zâ, Eyn** and **Qâf** are not used in pure Farsi words.

FA-script also has nine more characters and accents that decorate certain letters and modify its role and sound. These are: *Zir, Zebar, Pish, Sokun, Ye-ye-Kutâh, Tašdid*, Three sorts of so called *Tanvin* characters that we shall call *Do Zir, Do Zebar* and *Do Pish*.

Some aspects of the traditional FA-script can be characterized as follows:

- *Writing direction*: In FA-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*: FA-script lacks vowels. In principle the accents of the Arabic script (**Kasre, Zamme, Fathe, Sokun** (the latter is used to indicate non-vocal letters)) 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*¹. Several different sounds are represented by the same letter, for example, the occurrence of the Farsi-Arabic letter **Vâv** in the words: *to* (you), *ru* (on, above), *vali* (but), *rowšan* (bright). In these words,

¹These letters are distinct Arabic letters with distinct sounds and the problem we are describing here only concerns Farsi. Furthermore, their names are pronounced differently in Arabic.

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 FA-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.
- eFarsi alphabet is adequate for most dialects of Farsi.²
- eFarsi is primarily based on pronunciation³. In other words, whatever is written is pronounced and vice versa^{4 5}
- 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

²The examples in the report are written in the main stream accent of Iran.

³Following Aristoteles' words as to 'writing is the mirror of speech'

⁴For example, the eFarsi word *xâhar* is written as *xvâhr* in the FA-script. The main difference between the two is the occurrence of the letter *v*. As you see it does not appear in the eFarsi spelling. The reason for this is simply that the *v* is not pronounced and therefore ignored in the eFarsi version of the word.

⁵There is an exception to this rule, when the letters *n* and *b* appear immediately after each other, the *n* is usually pronounced as *m*. For example, *panbe* (cotten) is pronounced as *pam-bé*, *shanbe* (Saturday) as *sham-bé*, *donbe* (fat of sheep) as *dom-bé*. 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; words like *dom* (tail) and *som* (hoof) that are established Farsi words that originate from *donb* (tail) and *sonb* (hoof).

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

- 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.

â 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} ’

The names of these letters are:

â (*Â*), 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} (*Ŝe*), t (*Te*), u (*U*), v (*Ve*), w (*Ow*), x (*Xe*), y (*Ye*), z (*Ze*), \hat{z} (*Že*), ’ (*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 FA-alphabet. The pronunciation of the names follow the eFarsi pronunciation rules that will be explained later in this section.

Furthermore, there are two important diphtongs:

ow and *ey*

This alphabet contains the following vowels.

â a e i o u

â i u are long vowels and a e 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 sound of the eFarsi-letter.

â (Â)

- transliteration of certain occurrences of the Farsi-Arabic letter **Alef** such as *bâd* (wind), *bâdâm* (almond), *jodâ* (separate)
- transliteration of the **Alef e maddâh** (the Arabic name for the letter) or **Â-ye-bâ-kolâh** (the Persian name for 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 *Musâ* or *Esmâ'il*.
- 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 seldom written in FA-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), *andiâseh* (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'

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

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**, **Pish**) which is seldom written and it is the reader that has to determine the sign.

- some English words written in eFarsi: *kan* (can), *abzorb* (absorb), *asowsieyt* (associate)

b (Be)

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

c (Ce)

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

- transiteration of certain occurrences of the letter *eyn* in association with an implicit or explicit *Zir* vowel, for example, *eşq* (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 a Farsi 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), *raftar* (behaviour), *daftar* (office, book), *raft* (went), *fahmidan* (to understand)
- pronounced exactly like the 'f' in English

g (Ge)

- *g* is used for transliteration of the Farsi letter *Gâf*, for example, *agar* (if), *dâneşgâh* (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),

⁹Persian word for science is *danesh*.

rahbar (leader), *bahâr* (spring), *mâh* (moon), *kahkešân* (), *haft* (seven), *hašt* (eight), *noh* (nine), *Nuh* (Noah).

- pronounced as the 'h' in the English words 'hat', 'hot'.¹⁰

i (I)

- transliteration of the FA-letter **Ye** when used as a vowel, for example, *sib* (apple), *riz* (tiny), *niyat* (intention), *pâyiz* (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 Farsi 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* (guest).
- some English words written in eFarsi: *bin* (bean), *bin* (been), *bin* (bin), *sin* (sin), *sin* (seen), *Irân* (Iran), *binâyi* (vision).

j (Je)

- transliteration of the letter **Jim**, for example, *jân* (spirit), *xarj* (cost, expenditure), *jâ* (place), *injâ* (here), *ânjâ* (there), *juybâr* (water stream), *jâmedân*¹² (suitcase)
- pronounced as the 'j' in 'John', 'major'
- some English words written in eFarsi: *Jorj* (George), *Jân* (John), *jus* (juice)

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

¹²*jâmedân* is not used anymore, its more 'modern' form *camedân* is used instead.

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 FA-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

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 usually occurs at the end of words.¹³,

¹³Some 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

for example, *fe'lan* (for the time being), *ehthemâlan* (probably), *mozâfon elayh* (that to which is added)

- pronounced as the English 'n' in, for example, 'number', 'fan'

o (O)

- transliteration of the vowel **Pish Zamme**, 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 **Pish** (or **Zamme**), for example, *Otriš* (Austria), *ordak* (duck), *otobus* (Bus), *otomâtik* (automatic), *Orupâ* (Europe), *otu* (iron - for ironing)
- 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 Farsi letter **Pe**, for example, *pul* (money), *tup* (ball)
- pronounced as the English letter 'p'.

which heads an **Alef**. There are three cases:

- **Fathatan** (or **Tanvin e Fathe** or **Do Zabar**) is transliterated as *-an*, for example, *masalan* (for example), *eyman* (exactly, obviously), *aslan* (originally, inherently), *abadan* (never, not at all), *ettefâqan* (As it happens, coincidentally)
- **Dammatan** (or **Tanvin e Zamme Do Pish**) 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* ()

Tanvin only appears in Arabic words. Farsi words that may appear with **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*.

q (Qe)

- transliteration of the Arabic letter **Qâf** and the FA-letter **Qeyn**¹⁴, for example, *qamgin* (sad), *Qom* (Iraninan 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'

š (Ŝe)

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

¹⁴ **Qâf** and the FA-letter **Qeyn** are pronounced differently in Arabic, but in Farsi the pronunciations are in general very similar. In Farsi and some Arabic dialects, many occurrences of **Qâf** are pronounced as **Qeyn**. This is the motivation of transliterating both to *Qe* in eFarsi.

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'eratolma'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)¹⁵.
- pronounced as the 'oo' in 'tool', 'cool'

v (Ve)

- transliteration of some occurrences of the letter **Vâv**, for example, *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 'kh' in the names of some recent Iranian leaders 'Khomeini', 'Khamanei' and 'Khatami'.

y (Yâ)

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

¹⁵Very few Farsi words start with the letter *u*

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'

ž (Že)

- transliteration of the Farsi letter **Že**, for example,
- pronounced as the 'j' in French. In English, some occurrences of the letters 'g' and 's' are also pronounced like the Farsi ž, for example, the 's' in 'vision' or 'pleasure' and the second 'g' in 'Garage' (in some American pronounciations).
- 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'alle¹⁶* (teacher)
- transliteration of the **Hamze**. In Farsi, **Hamze** only appears in a medial position and indicates that there is a change of vowels.¹⁷ For example, *pâ'iz* (autumn), *miâ'i* (you come), *migu'im* (we say). In some cases, **Hamze** in Farsi words has been replaced by a **ye**. For *pâyiz* instead of *pâ'iz*, *miguyim* instead of *migu'im*.
- pronounced in the throat, just like the way the second 't' in 'that' or 'tt' in 'bottle' is pronounced in some English accents (Cockney or Mancunian). The sound generated by *Ist* is what linguists call a glottal stop which appears in many languages. Compare, for example, the

¹⁶The Farsi word for the Arabic *mo'alle* is *Âmuzgâr*.

¹⁷In Arabic the situation is more complicated. **Hamze** called **Hamza** is of two kinds: **Hamzatolvasl** and **Hamzatolqat'**. **Hamzatolvasl** aonly appears at the beginning and is dropped in certain situations. **Hamzatolqat'** can appear at the beginning, middle or the end of a word and can not be dropped. See [?] for more details.

sounds of the first and the second *a* in the English word *Adam*. The first *a* in Adam starts with a glottal stop.

Making *Ist* implicit

The occurrence of *Ist* between two vowels is for indicating a change in vowels, for example, *tavânâ'i* (power, strength). Explicit mention of *Ist* in the FA-script is important since vowels are implicit. In eFarsi, however, vowels are explicit and therefore it is possible to make the *Ist* implicit. That is, write *tavânâi* instead of *tavânâ'i*.

ey (Ey)

- transliteration of certain occurrences of the Farsi-Arabic letter **Ye**, for example, *peykar* (body), *peykân* (arrow), *peyâpey* (repeatedly), *key* (when)

ey pronounced as the *a* in the English 'take', *ei* in the English word 'eight' or the *ai* in 'Bahrain'; Also pronounced as the *ei* in the English word 'eight' or the *ai* in 'Bahrain' and 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), *nouruz* (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 Farsi, 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 diphtongs are presented for practical reasons.

3 eFarsi Conventions

3.1 *Ist*

Ist, the glottal stop, is denoted by a single quote or apostrophe (') and used for transliterating certain occurrences of the Farsi-Arabic characters **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 English¹⁸ but no written form is associated with it in the Latin alphabet. In Farsi, 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).¹⁹

¹⁸Glottal stops are common in English, for example, when pronouncing the expression 'uh-uh' meaning 'no' [7]. 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 [7]. 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.

¹⁹Here is 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).

3.2 Silent *h*

Some occurrences of the FA-letter *He-ye-Havvaz* are written to indicate that the word ends with a vowel (often *e* and sometimes *a*); *He-ye-Havvaz* in these cases 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 always written in order to keep the transliteration closer to the FA-script. For example, in *padideh* (phenomenon), *Zohreh* (Venus), *šâneh* (comb) and *parandeh* (bird).

In eFarsi we have chosen to drop all occurrences of the silent *h*. This is also true of the transliteration schemes Eurofarsi [1] and UniPers [2]. The above examples, are written as *padide*, *Zohre*, *šâne* and *parande*.

3.3 *Tašdid*

Tašdid in Persian and Arabic is a sign (similar to a small *w*) which is (supposed to be) placed on a consonant²⁰ 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 Farsi and English in that, in Farsi, 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*.

When *Tašdid* concerns the final letter of a word, the second instance of the doubled consonant is only pronounced if the word following it starts with a vowel. For example, in *Tâ in hadd soqut kard*, *hadd* is pronounced as *had* but in *Hadd o marzi nadârad*, it is pronounced as *had-d*.

²⁰In Farsi *Tašdid* is only applied to the letters *Pe*, *Ce*, *Re* and *Lâm*, for example, *tappe* (hill), *bacce* (child. kid). 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 plural, for example, *asrâr* (plural of *serr*), *hodud* (plural of *hadd*). This latter is not true of Arabic two-letter words without *Tašdid*, for example, the Arabic word *ab* (father) and its plural *abâ'*.

3.3.1 *Tašdid* on the Farsi-Arabic letter *Ye*

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

If there is a *Tašdid* on the FA-**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).

3.3.2 *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* (hinder). In these cases the pronunciation of the second occurrence of the letter is more apparent only when the word is followed vowel.

3.4 Ezâfe

Ezâfe (*e* or *ye*) is a prepositional morpheme used to indicate a relationship between two words according to the following format:

<mozâf> *e* <mozâfon elayh>
<mozâf> *ye* <mozâfon elayh>

Ezâfe in some cases is comparable to the English *'s* when used in possessive case and sometimes it functions as the French *du* or the English *of*. 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 dastur e zabân e Farsi e man*²¹ (my Persian grammar book).

²¹In *ketâb e dastur e zabân e Farsi e man* the words can be grouped as

In the traditional FA-script *Ezâfe* seldom appears in text. In eFarsi, however, it is written whenever it is pronounced. Because *Ezâfe* declares some sort of correspondence between two words, we have chosen to write it separately rather than concatenating it to the end of the *Mozâf* which is the case in some earlier transliteration schemes ([?]). Furthermore, writing it separately facilitates text analysis to some extent. *Ezâfe* appears in many different contexts (see the footnote²²

[[[ketâb e [dastur e zabân]] e Farsi] e man]

²²There are different sorts of *Ezâfe*:

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 shishe* (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), *Shahr e Tehran* (Tehran City), *ruz e šanbe* (day of Monday).
5. *Ezâfe ye Tašbihi* (Analogical *Ezâfe*) where analogy between the *Mozâf* and *Mozâfon Elayh* is drawn, for example, *mahd e zamin* (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), *gush e došman* (enemy's ear), *Dast e sabâ be zolf e saman šâne 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 Abollâh*), *Isâ ye Maryam* (Jesus son of Maria).
8. *Ezâfe ye Rabti* (Prepositional *Ezâfe*) where *Mozâf* is a primitive adverb or noun. For example, *bedun e to* (without you), *birun e dar* (outside the door), *zir e miz* (under the table)

for a more thorough description): 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 examples:

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 qermez (red color)
rang (color), *qermez* (red or crimson)

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)²³

As you have probably noticed, in *baraye* (for) the *e* is glued to the word *barây* which is never used without the *Ezâfe*. This is naturally an exception and one may include some other words in this exception list, for example, *bedun* (without) which is always used as *bedun e*.

Sometimes when *ezâfe* is used to connect noun and adjective, one can in principle, switch the position of the noun and the adjective and do without the *ezâfe*. For example, *ruz e âftâbi* (sunny day) which can be rephrased as *âftâbi ruz* (sunny day) are equivalent. This construct is quite common in some dialects of Farsi and also in Persian *poesi*.

²³Here are some 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 xoršid* (sunset), *joz' e kuchak* (small part), *farš 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 riyâzi e novin e man* (my modern maths book)

3.5 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 Farsi. There are many different sorts of combinations²⁴. In some European languages, such as German and Swedish constituents of a compound are usually concatenated²⁵, whereas in English compound words could be written in an open (spaced as separate words), hyphenated or a closed (set as one word) format[8]. The same sort of format is practiced in almost all Latin-based languages. For eFarsi, we have chosen to classify the formatting of compound terms as follows.

- *Gosaste* (separate): In this case the words of a compound term are written separately using a space as a delimiter. For example, *Emâràt 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).

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 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* style. For example, the compound term *gol e yax* (?) could after some persisting usage be written as *gol-e-yax* and as it is used more one may write it as *goleyax*. Naturally, finalization of such formatting requires decisions made by authorities such as the Iranian *Farhangestân e Zabân va Adab e Farsi*.

²⁴In [11] nine ways are suggested.

²⁵Examples: Swedish words *Aktiebolag* [*aktie* (share, equity) + *bolag* (company)](limited company), *regeringspartiet* [*regering* (government) + *partiet* (the party)] (the ruling political party)

Another useful point that should be mentioned is that writing words in the *Peyvaste* form may lead to some ambiguities or mispronunciations. For example, if a

- A prefix ends with the same letter as the the initial letter of the prefixed word, for exmaple, *hammihan* (fellow countryman).
- A suffix starts with the same letter as the final letter of the suffixed word, for example, *saxttar* (more difficult) from *saxt* (difficult).
- 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 wordform a diphthong. For example, suppose we use the diphthong *sh* instead of *sh*; the word *kushâ* could either be *kuâ* (A person who tries hard and does his best) or *kushâ* (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 and ambiguity.

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

3.5.1 What Words Are Combined

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

1. Two nouns, e.g., *kârxâne* [*kâr-xâne*] (factory), *shekarâb* [*shekar-âb*] (sweet drink made with suger and water), *Sâzmân e Melal* (United Nations), *Xalij e Fars* (Persian Gulf)
2. Two verbs²⁶ *keâkeâ* [*keâ-â-keâ*] (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'.)

²⁶This includes verb roots as well.

3. A noun and an adjective (*Sefidrud* (Sefidrud - white river)), *Siâhkuh* [*siâh-kuh*] (Siâhkuh - 'black mountain')
4. A numeral and a noun, for example, *cârpâ* [*câr*²⁷ (four) *pâ* (foot, leg)] (four-legged)
5. An adjective and a verb, for example, *zendebâd* [*zende-bâd*] (long live), *shâdbâsh* [*shâd-bâsh*] (congratulation)
6. 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)
7. A prefix and a word *begu* [*be-gu*] (say), *barnâme* [*bar-nâme*] (program), *pišraft* [*piš-raft*] (progress)
8. A word and a suffix, for example, *lâlezâr* [*lâle-zâr*] (tulip garden), *kârgar* [*kâr-gar*] (worker)
9. Adding an infix, for example, *keš-â-keš* (struggle)

Section 3.5.2 explains how these word combinations are constructed.

3.5.2 Building compound words

In this section we describe the various ways in which compound words can be created and suggest the appropriate format for writing them.

Simple concatenation

Simple concatenation of some nouns creates a compound word written in the *Peyvaste* format, for example, *bâqbân* [*bâq* (garden) *-bân* (-keeper)] (gardener), *serkeangabin* [*serke* (vinegar) *angabin* (honey)] (mixture of vinegar and honey - used as salad dressing)²⁸, *âbpâš* [*âb-pâš*] (watering-can)

²⁷ *câr* is a short form of *cahâr* (four)

²⁸ *Serkeangabin* has actually been transformed to *sekanjabin* 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 (6). 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*²⁹ (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 FA-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 write 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*.

Compound words using *Ezâfe*

Ezâfe is a common mechanism for constructing compound words in Farsi. Just like *o* in the previous section, *Ezâfe* works fine as a glueing mechanism 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).

However, dropping *Ezâfe* is a common phenomenon in Farsi. Once *Ezâfe* is removed from a construction, the glue is lost and it is important to discuss the writing rules for such cases. In fact, a number of different cases arise.

²⁹Normally used in the negative form as *bi kas o kâr* meaning a person who doesn't have any relative/roots in a community.

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 recommend that they should be concatenated to form a new word, that is 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 of a dictionary.

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 Farsi 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):

Pish az man o to, leyl o nahâri budast
Gardande falak niz, be kâri budast

Har jâ ke qadam nahi to bar ru ye zamin
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*³⁰. Similar examples are: *sib e sorx* (red apple) transformed into *sorx sib*. 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*³¹ (millstone) *sang-e-âsiâ* (millstone)

Examples for the case 2b: *golbarg* (petal) *barg e gol* (leaf of flower), *shâhpesar* (distinguished or exceptional son or boy) *pesar e shâh* (son of the king), *mehmân-xâne* (hotel, motel) *xâne ye mehmân* (guest's house), *gâvzabân* (borage) *zabân e gav* (cow's tongue), *narmafzâr* (software) *afzâr e narm* (soft tools), *xoshraftâr* (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 Farsi. Some common prefixes are: *mi*, *be*, *na*, *bar*. Some compound words containing these prefixes are: *miguyad* (is saying), *begu!* (say!), *nagu!*, *don't say*, *bargasht* (return, returned).

Adding Suffixes

All suffixes are simply concatenated to the end of a word and we suggest that the the suffix and the word should be written in the *Peyvaste* format

³⁰And most probably Xayyâm didn't mean to create a new word here.

³¹*Âsiâsang e zirin motaharrek nist, lâjeram tahammol e bâr e gerân konad*. Sa'di

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), *daneshmand danesh-mand* (scientist), *xeradmand xerad-mand* (wise).

Adding an infix

Infixes are not so common in Farsi, 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), *zanâsui* (Matrimony).

3.6 Exceptions

Ezâfe

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.

In and *Ân*

Vâzehâ ye eshâre (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), *ânce* (that which), *inhâ* (these), *ânhâ* (those), *inke* (this who), *ânke* (that who), *hamin* (this), *hamân* (that very), *conin* (such, such a one), *conân*³² (such, such a one), *inconin* (such a one), *ânconân* (such a one)³³, *hamconin* (just like this, also), *hamconân* (like before, as it has been, still), *candin* (several, so many), *candân* (so many, so much)³⁴, *intowr* (thus, in this way), *ântowr* (in that way), *ingune* (in this way), *ângune* (in that way), *ângâh* (then)

³²*conin* is also pronounced as *cenin* and *conân* is written as *cenân*. The *o* in *con* is pronounced as *e* in other similar constructions.

³³*Ânconân oftâd ke pâyaš šekast.*

³⁴*Candân ham gerân nabud* (It wasn't so expensive).

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). In the following cases, however, *be* is prefixed to the succeeding word and should be written in the *Peyvaste*-format:

- When it is written as *bed* in, and only in, the following words: *bedin*, *bedân*, *bedu*, *bedishân*.
- *Be* is sometimes used as the so called *Be ye Zinat* (redundant *be*) at the beginning of a verb or an infinitive. For example, *begoftam* (same as *goftam* (I said)), *begoftan* (to say), *beafkanad* (throws)³⁵, *beâ* (come)³⁶.

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 the object of a 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), *marâ* (short for *man râ* (me)) and *torâ* (you - as object of a sentence),

³⁵The common pronunciation of *beafkanad* is more like *biafkanad* with a shorter i-sound than usual

³⁶The common pronunciation of *beâ* is more like *biâ* with a shorter i-sound than usual

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* (), *hamcenân* (as usual), *hamin* (this one), *hamân* (that one), *hamânâ* (those ones), *bâ ham âmadim* (we came together).

3.7 Geographic Names, Proper Names and Trademarks

This transliteration convention does not apply to names of people and places. For example, although a better way of writing Geoffrey in eFarsi would be as *Jefri*, we still would write it as *Geoffrey* in order to preserve its spelling in the original language. In fact the most appropriate way of writing peoples names or registered trademarks 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.

3.8 Foreign words

For foreign words that are not geographic names or registered names, one of the following principle should be applied.

- If the original script is Latin-based, then the spelling in the original language would apply, for example, the English word 'president' would be written as *president* in eFarsi.
- If the original script is not based on Latin but there is a commonly used Latin-transliteration for it, then the spelling suggested by the Latin-transliteration should apply.
- Otherwise, a suitable eFarsi spelling, based on the pronunciation of the word in the original language, should be produced.

3.9 Some Modifications

3.10 Making eFarsi More Practical

In this section we introduce some diphthongs that are alternatives to some of the letters of the eFarsi alphabet. Computer users that for some reason do not have the letters \hat{a} , \hat{s} , \hat{z} or their equivalents in similar proposals [1, 2] would appreciate alternatives to these letters. It is mainly therefore we suggest the diphthongs *aa*, *sh* and *zh* below. We also introduce two more alternatives that are used traditionally, that is, *gh* and *ch* which are quite common in ranian names.

aa is an alternativ for \hat{a} . For example, *baayad* (must), *aayaa* (whether).³⁷

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

gh is an alternativ for *q*. For example, *Ghorve* instead of *Qorve*³⁸, *gholovv* (exageration).

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).

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 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.

³⁷Yet 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.

³⁸*Qorve* is a small city in western Iran.

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 *sĥ*. This would mean that the word *baâshâsĥ* (bright) would have to be written as *bashshaash*. It is very tempting to remove the first occurrence of the letter *h* since it is felt that it has no function there. It also turns out that by so doing esthetical aspects of the word improves. Compare *bachche* (child) with *bacche*, *oshshâq* (lovers) with *osshâq*, and *saghghez* (chewing gum) with *sagghez* which are alternatives to *bacce*, *oâshâq* 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 Farsi transliteration and transcription have suggested it independently. We considered the issue in 2002 but in April 2003 we learned that Dr. Sanamrad of IBM [9] 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, *fa'âl* (active). In eFarsi transliteration we have decided to reduce the double occurrence of *Ist* to only one since a double occurrence of the symbol is less appealing to the eye. For example, *fa'âl* is written as *fa'âl*.

3.11.3 *Ist* Between Vowels

As we mentioned earlier, the presence of *Ist* between vowels simply indicates a change in vowels. Although this would be important in the vowel-less FA-script, it is often redundant in eFarsi since the change of vowels is shown by the presence of the vowels themselves. It seems that we may be able to do away with the *Ist* sign that is placed between two succeeding vowels. Here are some examples mentioned in the text:

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

Compare these words 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 *oa* in these two words are pronounced differently, 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 vowel preceding it (*o*). In eFarsi, the sound of each vowel appearing in a word is not dependent on other letters in the word. A vowel is not used to modify the sound of another vowel.³⁹

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 have decided to introduce it in eFarsi. Here is a list of cases where capitalization should be applied.

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*
2. The first word of a sentence is capitalized. For example, *Yek sâ'at digar otobus miresad* (The bus arrives in an hour).
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

³⁹This is not to say that the pronunciation of two neighboring vowels do not affect each other. For example in section 4 where some common mistakes are discussed, we see that the pronunciation of *iâ* in some words has lead some transliterators to mistakenly use *iyâ* for achieving the sound effect that results when occurrence of *iâ* in some words is pronounced.

capitalized. For example, *Tehrân* (Tehran - capital of Iran), *Iâlât e Mot-tahed e Âmrîka* (United States of America), *Sâsmân e Melâl* (United Nations), *USA* (USA), *Ra'is e Koll e Qovâ* (Head of the the military forces - ground force, air force and the marines), *Majles e Showrâ ye Eslâmi* (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. Days of the week, months, and years are not capitalized. For example, *doshanbe* (Monday), *farvardin* (First month of the Iranian calender), *ut* (August), *sal e meymun* (year of the monkey)
6. Abbreviated titles are always capitalized. For example, *Âq. Rezâ Âshuri* (Mr. Rezâ Âshuri), *Dr. Ahmadi* (Dr. Ahmadi), *Xa. Simâ Shirâzi* (Mrs. Simâ Shirâzi), *Du. 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, *dâneşju ye Irâni* (Iranian student), *hame ye Irâniân* (all Iranians). This should also apply to other geographic units, for example, *bâzârâ ye Âsiâyî* (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 *Dâneshkade 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 *Sherkat e Naft e Irân* (Iranian Petroleum Company), *RI* as the abbreviation of *Râdyo 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*, *Duâsize 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 Farsi, ordinals (or ordinal numbers) are constructed by adding the suffix *-om*⁴⁰ or *-vvom* to a number. For example, *yekom* (first), *dovvom* (second), *sevvom* (third), *cahârom* (forth), and so on.⁴¹

We propose that these numbers be abbreviated by writing the numeral followed by a colon and 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 Farsi 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. 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.

⁴⁰When the number is not known, *cand* is used instead and the suffix *-om* is added to it: *candom*.

⁴¹As 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.

For these numbers we suggest the following abbreviations: 1min, 2min, 3min, 4min, ... or 1^{min} , 2^{min} , 3^{min} , 4^{min} , ...

3^{min} *Konferâns e Anjoman e Ânformâtik e Irân* (3rd Conference of the Informatics Society of Iran).

6. – 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 relationship).

3.12.3 Punctuation

eFarsi and FA-scripts, 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. [8]).

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 (*porsesh*) ends with a question mark (?). *Esmet chie?*
(What is your name)
3. An exclamation (*jomle ye ta'ajjobi*) ends with an exclamation mark (!).
Mage mishe târix e kešvarat râ balad nabâši (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 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 dârad: šanbe, yekšanbe, došanbe, sešanbe, cahâršanbe, panjšanbe, 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, Farvardin is 31 days and Mehr 30 days)
7. Quoting rules are as usual, for example, *Ârash goft: 'Midâni "omni-science" be Fârsi ci mishe?'* (Ârash said: 'Do you know what "omni-science" would be in Farsi?')

4 Common Mistakes

In this section we briefly discuss some of the common mistakes that occur when Farsi is transliterated. In particular, the letter *i* followed by other vowels *a*, *e*, *i*, *o* and *u* seems to be a major problem. The core of the problem is how to write a word that contains an I-sound immediately followed by another vowel. Usually, when pronouncing the sequence of the two vowels a Y-sound seems to appear in between. For example, when pronouncing *siâh* it seems that we are pronouncing the word *siyâh*. However, this is not the case, the Y-sound is simply the side-effect of pronouncing the sequence of vowels *i* and *â*.

Some more examples are provided in table ??.

Correct	English	Incorrect
<i>xiâl</i>	imagination	<i>xiyât</i> ⁴²
<i>piâde</i>	(on foot, pedestrian)	<i>piyâde</i>
<i>niâkân</i>		<i>niyâkân</i>
<i>siâh</i>	black	<i>siyâh</i>
<i>farhangiân</i>	academics	<i>farhangiyân</i>
<i>vidio</i>	video	<i>vidiyo</i>
<i>râdio</i>	radio	<i>râdiyo, râdyo</i>
<i>kiân</i>	name for boys	<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>

5 Discussion and Further Work

There are no standard orthographic principles for writing Farsi using the Latin script. There are a number of proposals but none of these proposals has

any official status. We hope that the orthographic principles we have sketched here eventually contributes towards a standard. But actual standardization is a complicated process beyond the scope of this article. The best we can hope for is that more and more people adopt eFarsi.

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 we have 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 Farsi text written in the FA-script and eFarsi.

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

5.1 Why Latin?

Our main reason 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, Farsi 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-sound. Here are some English words: *been*, *seen*, *deed*, *teacher*, *bit*, *rythm*, *clergy*. Actually this is too much of a simplification since things are much more complicated in English.

Anyhow, following the FA-writing system, we have decided to represent both the long and the short i-sounds with a single *i*.

- We have decided to write the *Ezâfe* separately, for example, *nur e xorâid* (light of the Sun), *pâ ye to* (your foot). This is also suggested in [1],

but [2] proposes concatenation of the *Ezâfe* to the end of the first word, for example, *nure xoršid* (light of the Sun), *pâye to* (your foot).

- The FA-letters *Qeyn* and *Qâf* have almost the same sounds in Farsi. 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 Farsi 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 Farsi-speaker 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*) should be sufficient for *Qeyn* and *Qâf* [1, 2, 10].

6 Appendix

Appendix A

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

Appendix B: More Examples

â: *âb* (water) *bârân* (rain) *bâdâm* (almond)
a: *abâ* (fear) *tars* (fear) *vahšat* (fear) *na* (no).
b: *bâmdâd* (morning) *abr* (abr) *šab* (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) *z*: *zarf* (depth) *pazhuheš* (research) *kuž*
()
zh: *zharf* (depth) *pazhuheš* (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) *meqhdâr* (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).
A poem by great Iranian poet Xayyâm

tâ chand hadis e panj o cahâr ey sâqi
moškel che yeki che sad hezâr ey sâqi xâkim hame, chang besâz

ey sâqi
bâdim hame, bâde biyâr ey sâqi

Appendix C: Examples of Compound Words

nâzparvarde [*nâz-parvarde*] (someone who is brought up with excessive love and affection - and probably spoiled);

mehmânnavâz [*mehmân-navâz*] (hospitable person);

xedmatkâr [*xedmat-kâr*] (servant, service personel);

xeradpiše [*xerad-piše*] (humble person);

gušenešîn [*guše-nesîn*] (withdrawn);

xodkâr [*xod-kâr*] (automatic);

xodxâh [*xod-xâh*] (selfish)
faryâdras [*faryâd-ras*] (helper);
barâdarzâde [*barâdar-zâde*] (niece, son or daughter of brother);
zaminlarze [*zamin-larze*] (earthquake);
âbpâs [*âb-pâs*] (watering-can);
farâmuşkâr [*farâmuš-kâr*] (absent-minded);
âtašafruz [*âtaš-afruz*] (person that starts a fire usually meaning a person that initiates heated discussions or loaded situations);
dowlatmard [*dowlat-mard*] (man of politics);
dastpoxt [*dast-poxt*] (cooking);
xošlebâs [*xoš-lebâs*] (someone that dresses well);
xartuxar [*xar tu xar*]⁴³ (higgledy-piggledy, chaotic)
jostoju [*jost o ju*, that is *jost va ju*] (search);
šostošu [*šost o šu*] (washing);
goftogu [*goft o gu*] (conversation);
kârobâr [*kâr o bâr*] (work);
âsopâs [*âs o pâs*] (hopeless);
seftosaxt [*seft o saxt*] (solid and hard);
didobâzdid [*did o bâzdid*] (visiting each other);
raftoâmad [*raft o âmad*] (commuting);
*kasokâr*⁴⁴ [*kas o kâr*] (relatives and close people);
*nâmonešân*⁴⁵ [*nâm o nešân*] (name and address); *xertopert* [*xert o pert*] (junk, pedlary);

Some of these constructions have gradually taken the form of a primitive single word. At the same time the *o* has been transformed to an *e*, for example, *josteju* for *jostoju*, *šostešu* for *šostošu*.

⁴³*xar* means donkey and *tu* means in.

⁴⁴Normally used in the negative form as *bi kasokâr* meaning a person who doesn't have any roots in a community.

⁴⁵Normally used in the negative form as *bi nâmonešân*

References

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