

Comparative Phonetic Study of Frequently Used Words in Iranian Farsi versus Tajik Farsi

Abdolazim Hakimi (Ph.D)

¹The Research Expert of Research Institute for Education,

² Department of English Language, Islamic Azad University- South Branch, Tehran, Iran
azimhakimi@yahoo.com

Abstract: This paper is a report on comparative phonetic study of the most frequently used words in contemporary Iranian and Tajik Farsi reference languages. It contains a brief historical overview of Farsi language with special focus on Iranian and Tajik dialects. After a discussion of reasons for differences in languages, this study attempts to identify the degree of phonetic similarities and differences of two language varieties. Corpus linguistic of this study was made of 3000 most frequently used words selected from an Iranian Farsi language database. These words were transcribed in both languages by using survey method. They were classified into 10 groups based on transcription differences. Study analysis concentrated on identical words with similar meanings in both languages. Phonetic differences of this group of words were studied in details. A systematic approach was used to classify phonemic alterations. This study provides its comparative analysis by using word samples. The outcome of the study showed that 18.23% of corpus linguistics was made of identical words with similar meanings and no phonetic differences. Another 77.37% of samples was made of identical words with similar meanings but with different phonetic. The existing differences between two Farsi language varieties are not so extensive to prevent communication or interaction of Iranian and Tajik Farsi speaking individuals. Considering the phoneme positions in two languages and their frequency in corpus linguistic, the most differences were observed in vowels. The differences of the two languages follow given phonetic tendency patterns. These phonetic tendencies symbolize each language.

[Abdolazim hakimi. **Comparative Phonetic Study of Frequently Used Words in Iranian Farsi versus Tajik Farsi**. Journal of American Science 2012;8(4):6 -16]. (ISSN: 1545-1003). <http://www.americanscience.org>. 2

Keywords: Farsi Reference Language, Most Frequently Used Words, Phonetic Differences, Consonant System, Vowel System, Iranian and Tajik Farsi Language Varieties

Introduction

Persian language has a long, extensive, and rich history. Different nations and tribes spreading across a vast area of Asian Continent have used this language for many centuries. Farsi has survived during its long history, sometimes spoken in a vast region and sometimes in a limited area.

Farsi is presently the official language of three countries, namely, Islamic Republic of Iran, Tajikistan, and Afghanistan. Cultural, political, social, and geographic developments in these three countries did not have extensive effect on Farsi language to limit interaction among these three nations and other Farsi speaking countries.

A comprehensive study in the history of Farsi language from early formation until today requires extensive work and many research programs to identify differences and similarities between Farsi language dialects. This research is limited to a comparative study of phonetic differences of frequently used Farsi vocabulary in Iran and Tajikistan.

This paper offers a short review of the historical background of Farsi language with a focus on Iranian and Tajik Farsi. It discusses the existing differences between the two language varieties and reasons behind them. This paper contains a

descriptive discussion of the phonemic differences of the two different dialects within the scope of the present study. The intention was to identify and present the phonetic disposition of the two language varieties.

Historical Background

The objective of this study was to identify the phonetic differences of the frequently used words in Iranian versus Tajik Farsi dialects. A brief historical review of these two dialects together with the changes and developments that resulted into the existing differences between two language varieties may be helpful in fulfilling this objective.

Farsi is a member of Persian language, which is a member of a larger Indo-Persian language family. This language family belongs to Indo-European group of languages, dubbed "Proto-Indo-European" language. (Sadeghi, 1978, p. 2)

The historical beginning of Persian languages and their separation from the main branch of Arian languages (Indo-Persian) dates back to second millennium BC. Past research and historical evidences indicate that the history of Persian languages covered three periods: ancient Persia, mid Persia, and contemporary Persia (Abolghasemi, 2003, p.16).

The discussions of Persian languages characteristics during the three historical periods are not within the scope of this paper. However, it is important to note that none of the Persian languages spoken in mid Persia survived except contemporary Farsi.

During the first and second centuries AH (*Anno Hegirae*) (7th and 8th centuries AD - *Anno Domini*), Farsi language took over dialects that were dominant in Western-Persia (Parthia). A vast area of Western Persia, including Khorasan, adapted Farsi language. Arabic was the scientific language of Iranians during this period. Non Moslem Iranians continued speaking mid Persian, Ashkani Pahlavi, Saghadi, and Kharazmi languages.

Iranian Moslems adapted Farsi Dari language with Arabic transcription during Saffarian Dynasty, starting from third century AH. (Abolghasemi, 2003, p. 37). Residence of mid Asia, covering Iran, Afghanistan, India, and Pakistan, spoke Farsi Dari. This language is now the contemporary Farsi used as the national language throughout Iranian territory.

Tajik Farsi has roots in Tajik literary language spoken in Mavar al-Nahr and Khorasan during fourth and fifth centuries AH. This language reached its pinnacle during Samanid Era. One thousand years of historical development and special influence of Russian revolution have resulted into the contemporary literary Tajik language.

Tajik Farsi has a long history but lacks written scientific or literary works. It was mostly a colloquial language. V. S. Rastorgueva had offered the most comprehensive classification of Tajik dialects. This classification grouped Tajik dialects into four groups, namely northern dialects, central dialects, southern dialects, and southeastern dialects (Rastorgueva, 1968, p. 56).

Historical Influences

A study into the history of Farsi language and the two Iranian and Tajik dialects reveals the differences and similarities between the two dialects. Farsi language has gained dissonance and heterogeneity under political, social, ecological, and economical influences compared to local Farsi dialects and language variations. Some of the differences and special characteristics found in different Farsi dialects stemmed from the influence of languages used by neighboring nations or languages dominant in the area prior to Farsi language adaptation.

The following is a brief discussion of the most important factors affecting dialect heterogeneity.

Cultural Separation

Assuming that Farsi dialects and varieties have roots in a single language, their present differences indicate that they have not gone through similar development path. Rather, they have developed as the result of or under the influence of interaction with different nations throughout their history. A historical review confirms this fact about Iranian and Tajik Farsi languages.

Shah Ismael Safavi defeated Mohammad Khan Sheybani (Khan-e Ozbak) in 916 AH. Dominance of Ozbak Turks in Mavara al-Nahr and Safavids ruling in Iran produced cultural and historical separation between Iran and Mavara al-Nahr. Adherence to different religious believes was the main reason for this separation. Safavid Kings strictly adhered to Shiite sect of Islam. Ozbak Khans prejudiced toward Sunnism.

Conflicts between Ozbak and Safavids exacerbated cultural and language separation of Iran and Mavara al-Nahr. The severity of conflicts in this part of Asia was so extensive that did not leave occasions for cultural unity (Ravaghi, 2005, p. 49).

The cultural and language separation between Iranians and Turks started from tenth century AD. The separation continued during Changizian and Taymoorian dynasties, who ruled the region for many centuries starting from 16th century AD. Russians entered Bukhara territory in second half of 19th century (1866 and 1867). This region became a part of Russian Federation after communists took over and Tsars were overthrown in Russia (1917). Russian influence in the area caused further separation of Iran and Tajik culture.

Shakoori Bokharaiee described this historical development: "During 1930s, not only many Russian, Greek, and Latin words entered Tajik language but also influenced colloquial Tajik language. Tajik language became more like Russian when Tajik words were transcribed in Russian alphabets." (Shakoori Bokharaiee, 2002, p. 64)

Language Influence

The most important reason for differences between Iranian Farsi and Tajik Farsi is the influence Mid Asian Farsi languages such as Kharazmi, Balkhi, Sokayee, and specially Saghadi have had on Tajik Farsi. The mixture of these Farsi varieties with Dari language and loaned Arabic words and transcription created a unique language in Mavara al-Nahr with specific words, structure, and phonetics. This language was dominant in that area and it could be one important reason for its heterogeneity with other Farsi language areas.

Oranski explained the differences in his statement: "Dari Farsi became dominant in Mavara al-Nahr in a short period in third century AH.

Zarafshan Valley residence and the neighboring areas who spoke Saghadi and those old Tokharestan regions who spoke different dialects of Balkhi, Sokaiee-Tokhariee gradually learned to speak Farsi. These areas became Farsi speaking areas after a period of speaking dual languages. Former languages of eastern Iran, namely Mavara al-Nahr and Tokharestan became gradually extinct. A part of this area adapted Farsi language and the other part adapted Turkish language. The area that adapted Tajik Farsi preserved certain phonetic and phonemic properties of their former languages." (Oranski, 1999, p, 104)

Cultural separation between Mid Iran and Mavara al-Nahr (starting from 10th century AD) and the influence of different Iranian dialects used then in Mavara al-Nahr area in Tajik Farsi have caused Tajik Farsi to become more distinctive from the Farsi language currently in use in Iran. Tajik Farsi now has different syntax, lexical, and phonetic forms when compare with Iranian Farsi.

Notwithstanding the differences, there are noticeable similarities between two Iranian and Tajik Farsi varieties in spite of historical, geographical, social, and cultural developments that have taken place over a long history. Linguistic scholars believe that if we compare Farsi scientific and literally works written during the last ten centuries, we could not identify significant localized linguistic differences such as word combination, grammar, or word construction (Bagheri, 1994, p. 111).

A Mavara al-Nahr scholar confirmed this conclusion by stating "The extensive use of Farsi language (Tajik) in a large area for a long period - from the end of third century AH (9th century AD) till the beginning of tenth century AH (16th century AD) - does not present significant localized differences in terms of etymology, grammar, and word construction" (Torsonov, 1989, p. 143).

The conclusions of studies conducted in this field indicated that Farsi Dari was the root of literary Iranian and Tajik Farsi (Oranski, 1999, p. 107). Similarity and homogeneity between two Farsi varieties are more than their divergence and heterogeneity. A new wave of political, cultural, social, and economic interactions is developing between Farsi speaking countries. These interactions may be instrumental in bringing the two Farsi dialects closer together.

Study Methodology

A brief introduction to historical background of Farsi language and its many branches and derivatives with a special focus on Iran and Tajik Farsi dialects, followed by a discussion of the historical and cultural reasons behind heterogeneities between two dialects,

can set the stage to explain the approach taken to identify and select frequently used words in Farsi usable language.

Farsi language consists of a wide variety of words including verbal, written, styling, and functional types developed and used over a long period. Therefore, linguistic research studies like this one should concentrate on documented factual and core data within a framework of a given corpus linguistic with adequate number of logical and suitable samples. The study should classify a huge volume of data into limited categories using criteria that fit the study objectives.

In consideration of study objectives and the requirement to limit the study scope, the contemporary versions of two Iranian and Tajik Farsi language varieties were selected as reference languages upon a synchronic dissection. The selected reference languages included millions of data. This study concentrated on most frequently used words to obtain a workable sample. This study used a fully developed measuring system to identify its scope and make its word selection with a focused dissection.

Initial investigations revealed that there were no previous studies on Tajik Farsi aimed at creating a linguistic database to help this study identify the frequently used words. Luckily, there was an extensive work on Iranian Farsi aimed at creating a Farsi linguistic database at national level. Human Sciences and Cultural Studies Institute, a scientific and research institution in Iran, had undertaken to carry out this project.

The objective of this project is to create a desirable and suitable corpus linguistic database. One feature of this database is an orderly and logical structure in spite of huge volume of diverse and extended linguistic data. Another feature of this database is rapid search and access capabilities to provide the required information on timely basis. The present database contains about one million words collected from various sources.

This section defines the most frequently used words, the significance of using this measuring criterion, and the approach used for selecting the frequently used words for this study.

Frequency in this study refers to the number of times a word is used in verbal or written form. Carroll used Foss views to explain the relationship between frequency of words and word access - the higher the frequency, the easier the access (1999, p. 118 & 119). In selection of words for verbal or written use, the most frequently used words become more active than the least frequently used ones.

The corpus linguistic of this study contains 3000 frequently used words selected from the existing database of Iranian Farsi language. The

frequency distribution of words in corpus verbal was the base for identification of core vocabulary. The frequency distribution of the top 2000 frequently used words in the corpus linguistics ended with a turning point marking a sudden decline. This point was a dividing point separating the core words from the rest. It was also a benchmark for separating top frequently used words.

The number of words and variety of colloquial texts in core area of written corpus is about 3000 words (Mac Carty and Carter, 2003, pp. 5-6). Considering the variety of written texts used for the selection of frequently used words, the number of 3000 was selected for the number of words in corpus linguistic of this study. The required analytical base for this study was created with the selection of this number of words from Iranian Farsi linguistic database using a declining distribution.

Phonetic Comparison

We examined the phonetic systems of Iranian and Tajik Farsi language dialects and compared word usage. The following represents the outcome of this comparison.

1. The first group of words started with a consonant. Farsi language has 23 consonants. Tajik has an additional consonant marked as /ɣ/ representing ħ. Therefore, Tajik Farsi has 24 consonants as follows: b, p, t, d, s, z, ċ, g, k, ĵ, q, ' , r, š, x, γ, ž, v, f, h, m, n, l, y.

2. The second group of words contained vowels that appeared anywhere in a word other than the first letter. Vowels appear in the middle or end of a word. Iranian Farsi has six single vowels, namely, i , a , e , u , o , â (Samareh, 1989, p. 39)

Tajik Farsi has six vowels; five of them are similar to Iranian vowels. Iranian Farsi has a unique vowel â, and Tajik Farsi has a unique ò vowel (Faiz, 1985, pp. 26-27). Although Iranian and Tajik Farsi languages have five similar vowels, their usage and frequency of use are not similar.

1. The third group of words contained diphthongs. Iranian Farsi has six digraphs or diphthongs, namely, ây, uy, oy, ay, ey, ow. These are diphthongs from phonetic viewpoint. They are combined sequence of a vowel and a consonant from phonemic viewpoint. Diphthongs do not have phonemic values in Iranian Farsi (Samareh, 1989, p. 122).

Diphthongs in Tajik Farsi are similar to diphthongs in Iranian Farsi from phonetic and phonemic viewpoint. The difference is that the number of diphthongs is four in Tajik Farsi versus six in Iranian Farsi. Four diphthongs in Tajik Farsi are uy, aw, ay, òy. Two of these diphthongs are similar to

Iranian diphthongs and the other two are different (Faiz, 1985, pp. 26-27 & Kalbasi, 1995, p. 54).

This study examined the sound structure within a single word and compared sounds and/or phonemes that created different sounds or produced different meanings. This study used a linear system of phonetic transcription for phonemic recording of the words under study in the two languages.

Phonetic Comparison

The selected words were first transcribed in Iranian Farsi. Next, these words were transcribed in Tajik Farsi based on different resources. A review of the resulting data revealed that not all selected words could be used to satisfy the study objective, i.e. comparison of the phonetic systems in Iranian and Tajik Farsi languages. This conclusion led to division of corpus linguistics into 10 groups according to word types.

Word Grouping

The following table provides a list of word groupings used to classify the study corpus linguistics.

Table 1: Grouping of Corpus Linguistics for Phonetic System Comparison

Group	Word Comparison	Identifier
1	Similar words, similar meaning, no phonetic difference	Similar-I
2	Similar words, similar meaning, different phonetic	Similar-II
3	Similar words with different meanings	Dissimilar-I
4	Different words with similar meanings	Dissimilar-II
5	Obsolete words in (one of the two) dialects	Obsolete
6	Foreign/borrowed words in (one of the two) dialects	Loanword
7	Incorrect division of two words as one entry	Compound
8	Incorrect division of a single word as one entry	Incomplete
9	Word repetition	Repetition
10	New words in Iranian Farsi (non-existent in Tajik)	New

Phoneme Grouping

The study sample was further classified by consonants, vowels, and diphthongs as additional analysis in comparing Iranian and Tajik Farsi languages.

1. Consonants

The following table lists samples of consonants according to their position in a word.

Table 2: Consonant comparison in Iranian and Tajik Languages

	Position	Word	Iranian Transcription	Tajik Transcription
A. Consonant to Consonant Alteration				
Consonant alteration from /q/ (ق) to /γ/ (غ)	Beginning	غذا	/qazâ/	/γizō/
	Middle	شغل	/šogl/	/šuryl/
	End	باغ	/bâq/	/bōγ/
Consonant alteration from /y/ to /ʔ/	Middle	دایره	/dâyere/	/dōʔira/
Consonant alteration from /b/ to /p/	End	اسب	/ʔasb/	/ʔasb/ /ʔasp/
Consonant alteration from /v/ to /b/	Middle	سرنوشت	/sarnevešt/	/sarnavišt/ /sarnabišt/
B. Consonant to /ø/ alteration				
Arbitrary alteration from /t/ to /ø/	End	عبارت	/ebârat/	/ʔibōra(t)/
Arbitrary alteration from /č/ to /ø/	Middle	بچه	/bačče/	/ba(č)ča/
C. Alteration of /ø/ to consonant				
Alteration of /ø/ to consonant /ʔ/ (a phoneme usually replacing "ع")	Middle	تاریخ	/târix/	/taʔrix/
Alteration of /ø/ to consonant /r/	Middle	علیرغم	/ʔalâraqm/	/ʔalōraqm/ /ʔalarraym/
Arbitrary alteration of /ø/ to consonant /t/	End	رتبه	/rotbe/	/rutba(t)/
Arbitrary alteration of /ø/ to consonant /m/	Middle	امید	/ʔomid/	/ʔum(m)ed/
	End	اعم	/ʔaʔam/	/ʔaʔam(m)/
Arbitrary alteration of /ø/ to consonant /k/	Middle	شکر	/šekar/	/šak(k)ar/
Arbitrary alteration of /ø/ to consonant /l/	Middle	طلا	/talâ/	/til(l)ō/
Arbitrary alteration of /ø/ to consonant /s/	End	تماس	/tamâs/	/tamōs(s)/
Arbitrary alteration of /ø/ to consonant /y/	Middle	ایمنی	/ʔimani/	/ʔe:mini:/ /ʔe:ymini/ /e:ymani/

2. Vowels

Iranian and Tajik Farsi Languages have five similar vowels. Each one of them has one dissimilar vowel; /â/ in Iranian Farsi and /ō/ in Tajik. Having similar vowels does not mean identical usage or same frequency. For example, vowel /o/ occurs with high frequency in Iranian Farsi but has very low frequency in Tajik. Comparatively, two allomorphs /e:/ and /a:/ occur with high frequency in Tajik (Kalbosi, 1995, p. 47).

The corpus linguistics was classified and studied from vowel phoneme viewpoint. The following is a discussion of the study outcomes.

No syllable in Iranian Farsi starts with a vowel, indifferent of its position at the beginning or middle of a word (Zemorodian, 2000, p. 55). When word transcription requires a syllable to start with vowel, a *hamzeh* with symbol /ʔ/ is added before the vowel. When a syllable in the middle of a word starts with a vowel, that syllable may start with a *hamzeh* /ʔ/ together with an intermediary phoneme /y/ based on phonetic circumstances. This represents one instance of pronunciation difference between Iranian and Tajik Farsi.

The following table contains samples that show such occurrences in the middle of words or between two syllables.

Table 3: Vowel Comparison in Iranian and Tajik Farsi Varieties

Vowel to Vowel Alteration	Position	Word	Iranian Transcription	Tajik Transcription
Alteration of Vowel /â/ to /a/	Middle	تابستان	/tâbestân/	/tōbistan/
	End	کالا	/kâlâ/	/kōlō/ /kōla/
Alteration of Vowel /e/ to /a/	Middle	شکر	/šekar/	/šak(k)ar/
	End	قصه	/qesse/	/qissa/
Alteration of Vowel /i/ to /a/	Middle	ریاست	/riyasat/	/rayōsat/
Alteration of Vowel /o/ to /a/	Middle	شمال	/šomâl/	/šamōl/ /šimōl/
Alteration of Vowel /i/ to /e/	Middle	زیبا	/zibâ/	/zebō/

Alteration of Vowel /e/ to /e:/	Middle	احتیاج	/ehtiyâj/	/e:htiyôj/
Alteration of Vowel /i/ to /e:/	Middle	ایران	/irân/	/e:rôn/
Alteration of Vowel /a/ to /i/	Middle	بایست	/bâyst/	/bô'ist/
Alteration of Vowel /e/ to /i/	Middle	اجرا	/ejrâ/	/i'jrô/
	End	آنچه	/ânçe/	/ônçi/
Alteration of Vowel /o/ to /i/	Middle	نقاط	/noqât/	/niqôl/
Alteration of Vowel /i/ to /i:/	End	الهی	/elâhi/	/ilôhi:/
Alteration of Vowel /a/ to /ô/	End	بالاخره	/bel'axare/	/bel'ôxira(t)/
Alteration of Vowel /â/ to /ô /	Middle	شاعر	/šâ'er/	/šô'ir/
	End	پیدا	/peydâ/	/paydô/
Alteration of Vowel /u/ to /ô /	Middle	میلیون	/milyun/	/milyôn/
Alteration of Vowel /a/ to /u/	Middle	جوان	/javân/	/juvôn/ /javôn/
Alteration of Vowel /e/ to /u/	Middle	چنان	/çenân/	/çunôn/
Alteration of Vowel /i/ to /u/	Middle	بیرون	/birun/	/burun/ /birôn/ /berun/
	Middle	قدرت	/qodrat/	/qudrat/
Alteration of Vowel /o/ to /u/	End	رادیو	/râdiyo/	/rôdiyu/
	Middle	صحبت	/sohbat/	/su:hbat/
Alteration of Vowel /o/ to /u:/	End	تنوع	/tanavvo/	/tanavvu:/
	Middle	گوشه	/guše/	/gu:ša/
Alteration of Vowel /u/ to /u:/	End	موضوع	/mowzu/	/mawzu:/
	Middle	پیرامون	/pirâmun/	/perôman/ /perômun/
Alteration of Vowel /ø/ to /e/	Middle	متر	/metr/	/meter/
Alteration of Vowel /ø/ to /ô/	Middle	انرژی	/enerži/	/inirziyô/

3. Diphthongs

Some Iranian Farsi references identified six double vowels (Samereh, 1989, p. 122). Other references identified minimum two diphthongs (Zemorodian, 2000, p. 52, Meshkot al Dimini, 2001, pp. 92-95).

The following table contains diphthong comparison in two Iranian and Tajik Farsi based on the corpus linguistics. The table offers different classifications depending on the position of vowels - occurring at the beginning, middle, or end of words.

Table 4: Vowel and Diphthong Alteration in Iranian and Tajik Farsi

Vowel to Diphthong Alteration and Vice Versa	Position	Word	Iranian Transcription	Tajik Transcription
Alteration of Vowel /o/ to Diphthong /aw/	Middle	رونق	/ronaq/	/rawnaq/
	End	جو	/jo/	/jaw/
Alteration of Diphthong /ow/ to Diphthong /aw/	Middle	دولت	/dowlat/	/dawlat/
	End	جلو	/jelow/	/jilaw/
Alteration of Diphthong /ow/ to Vowel /a/	Middle	اولویت	/o(w)laviyyat/	/alaviy(y)at/
Alteration of Vowel /i/ to Diphthong /ay/	Middle	نمی (بیشوند نفی ساز)	/nemi-/	/namay-/
Alteration of Diphthong /ey/ to Diphthong /ay/	Middle	والدین	/vâledeyn/	/vôlidayn/
Alteration of Diphthong /ow/ to Vowel /u/	Middle	موجب	/mowjeb/	/mujib/ /mu:jib/

Phoneme Alteration

Comparing words in two Farsi varieties included consonant and vowel phoneme alterations. Alteration sometimes involves alteration of a phoneme into a phoneme cluster. In other times, it

involves alteration of a phoneme cluster into another phoneme cluster. These alterations represent word usage differences, which may occur for variety of reasons. The next table provides samples of phoneme alteration into a phoneme cluster.

Table 4: Phoneme Alteration to Phoneme Cluster in Iranian and Tajik Farsi

Phoneme Alteration (Vowel with Consonant) or Alteration of a Phoneme Cluster with Another Phoneme Cluster	Position	Words	Iranian Transcription	Tajik Transcription
Arbitrary Alteration of Vowel /i/ with /nōk/	End	اساسی	/'asâsi/	/'asōsi:/, /'asōsnōk/
Alteration of Vowel /o/ with /yev/	Middle	اروپا	/'orupâ/	/'yevropō/
Alteration of Vowel /i/ with /ati/	End	زیادی	/'ziyâdi/	/'ziyōdati:/
Arbitrary Alteration of Vowel /e/ with /ina/	End	امروزه	/'emruze/	/'imru:zina/
Alteration of Vowel /ø/ with Consonant Cluster of /yō/ and /iyō /	End	انگلیس	/'englis/	/'anglisiyō/
Alteration of Phoneme Group /konân / to /garōn /	End	بازیکنان	/'bâzikonân/	/'bōzi:garōn/

Language Study

This section reviews the property of the information obtained from word groups, their frequencies, and their percentages based on phoneme differences in Iranian and Tajik Farsi. This review includes phonetic differences, the number of occurrences within a word, and the general phonetic tendencies.

Word Grouping

Study sample includes 3000 frequently used words in Iranian Farsi. Word grouping of corpus linguistics is explained below.

- Group 1 identified as *similar-I*: This group of words represents the similarity between two Farsi varieties. It significantly proves the similarity between Iranian and Tajik languages. This study revealed 542 words, or 18.23% of the study sample, pronounced similarly in both languages.
- Group 2 identified as *similar-II*: This is an important group in this study as it identified the phonetic differences in the two languages and their classifications. This group contained 2321 out of 3000 selected words, or 77.37% of the study sample.
- Group 3 identified as *dissimilar-I*: The number of words in this group is small. Nevertheless, it was important to separate them into an individual group for this study.
- Group 4 identified as *dissimilar-II*: The number of words in this group was 26 representing a small group.
- Group 5 identified as *Obsolete*: The study sample did not include frequently used words in Tajik that are obsolete in Iranian Farsi. It only included frequently used words in Iranian Farsi that are obsolete in Tajik language. The number of these words was minuscule.
- Group 6 identified as *Non-Farsi/Loanword* in one of the two language varieties: These words are borrowed from other languages based on various political, social, cultural, economic, or geographic circumstances. Two Farsi language varieties do not have similar loanwords. The sample study did not include frequently used loanwords in Tajik Farsi. The study sample had 45 loanwords. As the result of

Arabic language influence on Iranian Farsi, many Arabic words with small alterations and new identities have entered in Iranian Farsi language. These words are loanwords in this study.

7. Group 7 included *Compound*, *Incomplete*, and *Repetition* words. These three word types may have entered into study sample because of software problems or unknown reasons. The total number of these words was small in the study sample. They counted as 14, 14, and 35 words, respectively. These words were included in the study in spite of the small numbers.

8. Group 8 identified as *New* words: New words are recently constructed and entered in Iranian Farsi. They do not have equivalent in Tajik Farsi and, therefore, are not understandable by Tajiks. The number of new words was 19 in the study sample.

From this grouping and statistical data it became clear that this comparative study should focus on *Similar-I* and *Similar-II*, making 18.23% and 77.37% of the study sample, respectively. The number of words in these two groups could be a conclusive proof for closeness and similarity of these two languages from phoneme viewpoint at word level.

The high concentration of words in *similar II* group indicated that the main difference between two Farsi language varieties from phonetic viewpoint was marked by the differences observed in phonemic pronunciation at word level in specific phonetic circumstances. The frequency and type of these differences in different phonetic circumstances show linguistic tendencies of each Farsi dialect.

Language Comparison

This section provides a comparative study of phoneme grouping and their classifications based on consonants, vowels, diphthongs, and phonemic clusters. The comparison included analysis of phonetic differences and tendencies in each Farsi language variety.

A. Consonants

The important differences between Iranian and Tajik Farsi varieties resulted from comparing consonants in corpus linguistics are as follows:

1. Consonants exist in both Farsi language varieties from the phonemic distribution viewpoint. They appear at beginning, middle, and end of words
2. The frequencies of consonant positions are not the same in two language varieties. Consonants positions in words depend on the nature of the language.
3. The most important difference between Iranian and Tajik Farsi dialects is the different usage of two consonants ق and غ. Iranian Farsi does not differentiate between consonants ق and غ. They are pronounced similarly in Iranian Farsi with /q/ sound. In Tajik Farsi, however, consonants ق and غ are pronounced differently in conformance with Arabic pronunciation rules. The frequency of consonant alteration from /q/ (ق) to /ɣ/ (غ) in Tajik Farsi was 41 words in the study sample, regardless of their position in the word, being at the beginning, middle, or end.
4. In Tajik Farsi, a *hamzeh* ' / is inserted in phoneme transcription of a mid-syllable that starts with a vowel. In Iranian Farsi, an intermediary phoneme /y/ is sometimes used instead of *hamzeh* ' / depending on phonetic disposition. The number of alterations from consonant ' / to intermediary phoneme /y/ in Iranian Farsi was 56 words in the study sample. This number was high occurrence.
5. One notable alteration is alteration of consonant /b/ in Iranian Farsi to consonant /p/ in Tajik Farsi. A similar alteration is alteration of consonant /v/ in Iranian Farsi to consonant /b/ in Tajik Farsi. These alterations regularly occur in Tajik Farsi but do not occur in Iranian Farsi. The number of these phonetic alterations was very limited in the study sample.
6. Another difference was addition of consonant phoneme to pronunciation chain of Tajik Farsi. The highest number of consonant phoneme alteration within the framework of additional process belongs to arbitrary consonant alteration /ø/ in Iranian Farsi to consonant /t/ in Tajik Farsi. The number of this type of alteration was 32 in corpus linguistics. Similar consonant phoneme alterations included alteration of consonant /ø/ in Iranian Farsi to other consonants in Tajik Farsi, such as /m/, /r/, /k/, /l/, /s/, and /y/. There were very few alterations of this type in the study sample.
7. Another difference was elimination of consonant phoneme from pronunciation chain in Tajik Farsi within elimination process framework. This type of alteration included arbitrary consonant alteration /t/ in Iranian Farsi to consonant /ø/ in Tajik Farsi. Another example of this type of alteration was phoneme consonant alteration /č/ in Iranian Farsi to /ø/ consonant in Tajik Farsi. There are very few alterations of this type in the study sample.
8. Pronunciation of phoneme /k/ was another notable difference. Iranian Farsi speaking population

pronounce phoneme /k/ further back compared to Tajik Farsi speaking. This /k/ is an allomorph because different pronunciation did not represent phonemic difference.

9. Tajiks pronounce consonant /v/ in Arabic words with both lips like a /w/. This type of pronunciation was not common. Therefore, this study considered only phoneme /v/ pronounced by a lip and teeth. Other pronunciations of this phoneme in form of /w/ counted as allomorph.

B. Vowels

Comparison of two Farsi language varieties within corpus linguistics revealed notable points as discussed below. It is necessary to point out that phonemic alteration at word level may occur several times.

1. As a rule, vowels do not appear in all parts of Farsi words from the phonemic distribution viewpoint. Vowels appear only in the middle or at the end of words. In Farsi words that include one of written vowel symbols, a *hamzeh* or ' / is pronounced before that vowel. Phoneme transcription of words clearly shows this occurrence.

2. The frequency of vowels in different parts of a word is not similar in both languages. It means that some vowels have higher frequency than others do. For example, the frequency occurrence of vowel /a/ at the end of Farsi words is very low. The frequency occurrence of vowel /e/ at the end of Farsi words is very high. These observations point to characteristics of one language or dialect variety in applying specific vowels.

3. The most notable difference between two Iranian and Tajik languages is the frequency of alteration from vowel /â/ to /ô/ regardless of whether it occurs in the middle or at the end of a word. The number of this vowel alteration was 1451 out of 3000 words in the study sample. This represents the most frequent vocal characteristics of Tajik Farsi, which can differentiate Tajik from other Farsi varieties.

4. Another frequent vowel alteration in Tajik Farsi was alteration from /e/ to /i/. The number of this phonemic alteration was 726 words in the study sample. This type represents another characteristic of Farsi Tajik, which could be considered as an indicative characteristic of this Farsi dialect.

5. Vowel alteration from /e/ to /a/ is another notable difference of Tajik Farsi. This phonetic alteration had 547 occurrences in study sample representing the most frequent vocal change in Tajik Farsi language. As a rule, words ending with unpronounced (◌) /e/ in Iranian Farsi (such as چشمه, خانه, لانه) is replaced by /a/ sound in Farsi Tajik.

6. Alteration of vowel /o/ to /u/ in Tajik dialect was another notable difference. The number of occurrence

of this type was 450 in the study sample. This type of alteration was specific to Farsi Tajik.

7. Alteration of vowel /i/ to /i:/ appeared at the end of words in both Iranian and Tajik Farsi. The number of occurrence of this sound change was 410 in the study sample. This type of alteration is highly frequent in Farsi Tajik.

8. Alteration of vowel /u/ to /u:/ appeared only in Tajik Farsi words. This phonetic alteration is a sound change and had 91 occurrences in corpus linguistics.

9. Alteration of vowel /i/ to /e/ appeared in the middle or at the end of Tajik Farsi words. The frequency occurrence of this type was 74. It is one of phonetic alterations in Tajik Farsi.

10. Alteration of vowel /a/ to /i/ is another difference between two Farsi varieties. The frequency occurrence of this phonemic alteration was 37.

Vowel alteration from /e/ to /e:/

Vowel alteration from /a/ to /u/

Vowel alteration from /o/ to /i/

Vowel alteration from /o/ to /a/

Vowel alteration from /u/ to /a/

Vowel alteration from /o/ to /ō/

Vowel alteration from /u/ to /ø/

14. In rare cases, there are phonemic changes converting a vowel to a diphthong. Cases of vowel alteration in Iranian Farsi to diphthong in Tajik Farsi are as follows, listed by frequency:

Alteration of vowel /o/ to diphthong /aw/

Alteration of vowel /i/ to diphthong /ay/, and

Alteration of vowel /i/ to diphthong /ey/

15. Another difference between the two languages is the alteration of indeterminate vowel in Iranian Farsi to phonemic cluster in Tajik Farsi. Cases of phonemic change from Iranian Farsi to phonemic cluster in Tajik Farsi are:

Alteration of vowel /i/ to phonemic cluster /ati/, and

Alteration of vowel /e/ to phonemic cluster /ina/

C. Diphthongs

Comparative study of diphthongs in Iranian and Tajik Farsi dialects using study corpus linguistics showed the following points:

1. The most notable difference in studying diphthongs in two Farsi dialects is the alteration of /ow/ to /aw/ in Tajik Farsi. The frequency of this type of sound change was 47 in the study sample. This is the most distinguished difference in diphthongs that differentiate Tajik Farsi from Iranian Farsi.

2. A special characteristic of the two languages is the alteration of diphthong /ey/ in Iranian Farsi to /ay/ in Tajik Farsi.

3. There are cases where a diphthong in Iranian Farsi is altered to a vowel in Tajik Farsi. These cases

11. Alteration of vowel /i/ to /e:/ had high frequency in Tajik Farsi. Iranian Farsi does not have specific phoneme for this alteration. The reason is that Iranian Farsi does not have a recognizable vocal strain for this alteration and this vowel is an allomorph. The number of words with this type of vowel was 23 in Tajik Farsi.

12. Alteration of vowel /o/ to /u:/ is recognizable in Tajik Farsi. This phonemic alteration is specific to Tajik Farsi and had a frequency of 20 words in corpus linguistics.

13. There are other vowel differences in two Farsi language varieties with very low frequency. Nevertheless, they were still represented in the study sample. These phonetic alterations from Iranian Farsi to Tajik Farsi, listed by frequency, are as follows

Vowel alteration from /o/ to /a/

Vowel alteration from /â/ to /a/

Vowel alteration from /i/ to /a/

Vowel alteration from /u/ to /o/

Vowel alteration from /o/ to /u/

Vowel alteration from /i/ to /u/

Vowel alteration from /ø/ to /e/

that have very low frequencies are: alteration of diphthong /ow/ to vowel /a/, alteration of diphthong /ow/ to vowel /u/, and in some case alteration of diphthong /ow/ to vowels /u:/ and /o/.

D. Units Larger than Phoneme

Comparative study of units larger than phoneme revealed a few samples such as alteration of phonemic group /yō/ to /iyō/ and /konân/ to /garōn/ in Tajik Farsi.

Conclusion

Comparative phonetic study of frequently used words in Iranian and Tajik Farsi languages included: a) grouping of corpus linguistics into 10 groups, b) identifying word characteristics and specifications, c) measuring usage frequency and significance of each group, and d) examination of study sample based on consonants, vowels, diphthongs, and phonemic clusters. The following are the outcome of this comparative study:

1. Corpus linguistics contained 18.23% identical words with similar meanings and no phonetic differences (group 1) in both languages. Identical words with similar meanings but with slight phonetic differences (group 2) made up 77.37% of the study sample. The number of words in these two groups proves the connectivity and similarity of these two languages from phonemic viewpoint at word level. It also shows that the main difference between two

language varieties from phonetic viewpoint is the differences observed in pronunciation of certain phonemes at word level within given phonetic dispositions. The number of words in these two groups also indicates that the differences between two language varieties do not hinder communication and interaction between two nations.

2. Comparative study of consonants and diphthongs in two language varieties indicated that there was no difference in the possibility of occurrence of each phoneme at word level, regardless of appearing at the beginning, middle, or end of a word. The highest difference was for vowels at word level.

3. Frequency of vowels is not identical in two languages. For example, the frequency of vowel /o/ is very high in Iranian Farsi and very low in Tajik Farsi. Two allomorph /e:/ and /a:/ appear with high frequency in Tajik Farsi.

4. Phonemic differences in study corpus linguistics follow a distinctive pattern from phonetic viewpoint, especially in case of vowels. These patterns could act as unique identifiers of one language versus the other. One example is alteration of vowel /â/ and /e/ in Iranian Farsi to vowel /ô/ and /a/ in Tajik Farsi. Another example is alteration of vowel /e/ in Iranian Farsi to vowel /i/ in Tajik Farsi. One other example is alteration of diphthong /ow/ in Iranian Farsi to diphthong /aw/ in Tajik Farsi.

This study used a given corpus linguistic made of the most frequently used words collected from an Iranian Farsi database. This corpus linguistic was used as a reference for identifying phonetic differences between Iranian Farsi and Tajik Farsi. The approach used in this study was a new way of conducting comparative language studies to be adapted for future research projects. Collection of most frequently used words in target languages, their phonemic transcriptions, identifying similarities and differences of languages, together with analysis of findings could determine the degree of connectivity and phonetic dispositions of two languages or dialects.

Corresponding Author:

Abdolazim Hakimi

¹The Research Expert of Research Institute for Education,

² Department of English Language, Islamic Azad University- South Branch, Tehran, Iran

azimhakimi@yahoo.com

Tel:+98216640154

Mobile:+ 09122190361

azimhakimi@yahoo.com

References

1. Abolghasemi, M. "Farsi Language and Its Story", 2nd Edition, Hirmand Publishing, 2003.
2. Abolghasemi, M. "History of Farsi Language", Tehran: Organization for Study and Publication of Human Science University Books (Samt Publishing), 1994.
3. Asher, R. E. and Henderson, E. J. A. (ed.), "Towards a History of Phonetics", Edinburgh University press, Edinburgh, 1981.
4. Asi, M. "Farsi Language Database in Internet", *Research Journal*, No. 2, Human Sciences and Cultural Studies Institute, 2005.
5. Bagheri, M. "History of Farsi Language", 1st Edition, Ghatreh Publishing, 1994.
6. Bakalla, M. H. "Ancient Arab and Muslim Phoneticians: An Appraisal of Their Contribution to Phonetics", in Hollien, H and Hollien, P. (eds) Current Issues in the Phonetic Sciences. Proceedings of the IPS-77 Congress, Miami Beach, Florida, 17-19th December 1977. Benjamins, Amsterdam, part I:3-II, 1977.
7. Carroll, D. W. "Psychology of Language", London: Brooks/Cole, 2008.
8. Christensen, A. "Contributions á la dialectologie iranienne", v. 1-11, københavn, 1933.
9. Feyz Off, M. "Talafoz-e Adabi Tajiky", Dushanbe, 1985.
10. Frejman, A. A. "Xorezniskij jazyk", Moscow-Leningrad, 1951.
11. Haghshenas, A. M. "Phonology", Tehran: Agah Publishing, 1977.
12. Haghshenas, A. M., et al. "Farsi Language 1 for First Year High School", 13th Edition, Office of Planning and Editing Educational Books, Iran Educational Books Publishing, 2009.
13. Henning, W. B. "Acta Iranica", 14-15, Belgium, 1977.
14. Kalbasi, A. "Tajik and Iranian Farsi: A Comparative Study", Tehran: Ministry of Foreign Affairs Publishing, 1995.
15. Kozlov, G. I. "Različija između tadžikskim i persidskim jazykami", M., 1949.
16. Lazard, G., "La langue des Plus Anciens Monuments de la Prose Persane", Paris, 1963.
17. Lazard, G., "Caracteres distinctifs de la langue tadjik", *Bulletin de la Societe de linguistique de Paris*, t.52, fasc. 1, Paris 1957.
18. Le coq, P. "La langue des inscriptions achéménides", *Acta Iranica, Ière série, commémoration Cyrus*, Hommage Universel, II, 1974.
19. Meshkatoddini, M. "Phonetic Construct of Language", 2nd Edition, Mashhad: Ferdosi University Publishing, 1991.
20. Natel Khanlari, P. "History of Farsi Language", in 2 Volumes, 3rd Edition, Tehran: Nashr No, 1987.

21. Nazarzadeh, S. , Sanginov, A., Karimov, S., and Soltan, M. H. "Interpretational Dictionary of Tajik Language", Tajikistan Sciences Academy, Roodaki Language and Literature Institute, 2006.
22. Oranski, Y. M. "Iranian Languages", Translated by Sadegh, A. A., Tehran: Sokhan Publishing, 1999.
23. "Основы иранского языкознания Новоиранские языки", Восточная подгруппа. Москва, 1987.
24. Perry, J.R., "Comparative Perspective on Language Planning in Iran and Tajikestan", Language and Society in the Middle East and North Africa, Ed by Yasir Suleiman, Britain: Curzon Press, 1999.
25. Rastorgueva, V. S. "Mid Asia Farsi Language Grammer", Translated by Shadan, V., Iranian Culture Foundation Publishing, 1968.
26. Rastorgueva, V. S., (ed.), "Novoiranskiye yazyki, vostochnaya gruppya", Moskva, 1987.
27. Ravaghi, A. "Mavara al Nahr Farsi Language", Hermes Publishing, 2005.
28. Riazul Islam, "Indo-Persian Relations", Lahore, 1970.
29. Rozenfel'd, A. Z. "Tadžiksko-persidskue jazykovye otnošenija" /po materialam leksiki/učenje zapisie LGU 294\=serija vostokovedčeskix nauk, 12-42, 1961.
30. Sadeghi, A. A. "Farsi Language Development", Tehran: Iran Azad University Publishing, 1996.
31. Sadeghi, A. A. "Historical Differences of Iranian Farsi versus Tajik Farsi", *Kaihan Farhangi (Tajik Cultural Issue)*, V. 10, No. 103, Tehran: 1993.
32. Sadeghi, A. A. "Reference Language", Poorgavadi, N. (Ed.), *Selected Nashr Danesh Articles on Farsi Language*, Tehran: University Publishing Center, 1996.
33. Samareh, Y. "Farsi Language Phonology", 2nd Edition, University Publishing Center, 1989.
34. Shakoori Bokharaee, M. , et. al. "Tajik Farsi Dictionary", Translated and Edited by Shojaee, M., 2nd V., Tehran, Farhang Moaser Publishin, 2006.
35. Shakoori Bokharaee, M. "Humanism, Education, and National Language", Translated by Behrouzian, A., Dushanbe: Islamic Republic of Iran Cultural Center Publishing in Tajikistan, 2002.
36. Shojaee, M., *Farsi Language Problems in Tajikistan*, V. 9 , N. 56, Tehran: Danesh Publishing, 1989.
37. Torsonov, A. "Ehyayeh Ajam", Saeedov. F. (Ed.), Dushanbe: Erphan Publishing, 2000.
38. Vennemann, Theo, and peter Ladefoged. 1971. "Phonetic Features and Phonological Features." Working papers in phonetics 21.13-24, University of California, Los Angeles.
39. Zomorrodian, R. "Guideline for Collect and Description of Dialects", Mashhad: Ferdosi University Publishing, 2000.
40. Amonova, F. R, "Imennoe affiksāl'noe slovoobrazovanne v sovremennom tadžikskom jazyke (v sopostavlenii s persidskim)", Dušanbe, 1979.
41. Айнӣ, Садриддин. "Луғати Нимтафзилии Тоҷикӣ", Нашриёти Ирфон, Душанбе, 1976.
42. Галунов, Р.А., "эстеляхаты, Записки ИВАН", т.І, изд-во АН СССР, 1932.
43. Гаприндашвили, Ш.Г., "Гиунашвили, Дж". Ш., Фонетика персидского языка. Тбилиси: Мецниереба, 1964.
44. Жуковский, В.А., "Материалы для изучения персидских наречий", ч.І.Пр,1922.
45. Поляков, К.И., "Персидская фонетика: опыт системного исследования", Москва: наука, 1988.
46. Файзов Маърам. "Гласные звуки таджикского литературного языка" (акустический и количественный анализ) иадетебстбо Душанбе-1992.185с.
47. Файзов Маърам. "Таджикское литературное произношение", Душанбе-1985.

2/12/2012

2/12/201