

A Historical Study of the Persian Vowel System*

Elham Rohany Rahbar
University of Toronto

1. Introduction

It is widely believed that the Middle Persian vowel system was quantitative (e.g. Salemann, 1930; Rastergueva, 1969; Windfuhr, 1979; Pisowicz, 1985). This vowel system changed over time to result in its current main dialects: Dari (spoken in Afghanistan), Modern Persian (spoken in Iran), and Tajik (spoken in Tajikistan).

The goal of this paper is to account for the development of the Middle Persian vowel system to these three dialects. The framework within which I present my analysis is modified contrastive specification (Dresher, Piggott and Rice, 1994). In particular, I follow the view that contrastive specification is the result of ordering features into a contrastive hierarchy (Dresher, 2003a, 2003b, 2003c; Dresher and Xi, 2005).

This paper is organized as follows: Section 2 contains background on the vowel systems of Dari, Modern Persian, and Tajik. Section 3 suggests historical changes in terms of contrastive features in the inventories of these dialects. Section 4 concludes.

2. The Modern Persian, Dari, and Tajik vowel systems

In the evolution of the Persian language, the middle era started in 331 BC and ended by the Arab conquest of Iran in 652 AD, which is considered the starting point of the modern era (Bahar, 1942; Natel Khanlari, 1987; among others).

Dari, Modern Persian, and Tajik, the main dialects of Persian in the present time, are mutually intelligible. Some differences exist in their sound systems. The Middle Persian vowel system and the vowel systems of the three dialects are given in Figure 1.¹

Dari is the closest dialect to Middle Persian; Modern Persian shows more changes than Dari compared to Middle Persian; Tajik is thought to have been under the influence of Turkic and has its own characteristics. The Middle Persian vowel system was quantity-based as mentioned above. Dari is the only dialect which preserves quantity. The only change Dari shows since Middle Persian is lowering of *i* and *u* to *e* and *o*. It is generally agreed that the Modern Persian vowel system is qualitative (Samareh, 1977, 1985; Pisowicz, 1985; Meshkatod Dini, 2000; among others). Modern Persian underwent three changes: fronting of *a* and backing of *ā*; merging of *ē* and *ō* with *ī* and *ū*; and lowering of *i* and *u* to *e* and *o* (e.g. Pisowicz, 1985).

* Special thanks to Elan Dresher for supervising this work and for his invaluable advice, guidance, and support. I am also grateful to Keren Rice for many thoughtful suggestions on drafts of this paper. Thank you also to Yoonjung Kang for her insightful discussions on the later version of this paper. Any errors in this paper are mine.

¹ The system of Tajik is from Perry (2005), and the one of Dari from Pisowicz (1985). In Pisowicz, the low vowels are on the top and the high vowels in the bottom. I change the arrangement to conform to the other given inventories.

change that occurred in Tajik, is rounding of the long back low vowel in the direction of *o*. In Tajik, *e* is a continuation of \bar{e} . The vowel / \bar{u} / (which will be discussed below) has different sources according to Perry: it can be a continuation of \bar{o} (e.g., Tajik *rūz* ‘day’ corresponds to Dari *rōz*). Similarly, in Uzbek and Turkic loanwords (e.g., *kūmak* ‘help’ which is *komak* in Modern Persian). It can also be the result of the lowering of short /u/ before /h/ or glottal stop (e.g., *kūhna*). The position of / \bar{u} / is under Uzbek influence, as Perry mentions. Phonetically, it is between [u] and [y]. This vowel is phonemic only in Northern dialects. In Central and Southern dialects it is generally /u/. An example is /*kūhna*/ or /*kuhna*/ ‘old’ which is *kohne* in Modern Persian.² The vowels /i/ and /u/ are the continuation of \bar{i} and \bar{u} and they also represent short /i/ and /u/ (e.g., *dil* ‘heart’, *but* ‘idol’³). It is important, as Perry mentions, not to confuse /i/ and /e/ in Tajik and Modern Persian because in fact they are reversed; for instance, *bedil* in Tajik is equivalent to *bidel* in Modern Persian meaning ‘heartless’.

The merging of Middle Persian \bar{i} and \bar{u} with their short counterparts, and also the merging of \bar{a} with *o* cause some homonymy in Tajik which is not observed in Modern Persian. For example: the Tajik words *bi:no* ‘sighted’ and *binō*⁴ ‘building’; Persian *boridan* ‘to cut’ and *baridan* ‘to rain’ as Tajik *boridan*. Recall that in Modern Persian merging of the former \bar{e} and \bar{i} caused some homonymy as in *fīr* ‘lion, milk’ which were Middle Persian *fēr* ‘lion’ and *fīr* ‘milk’ (and which are still the same as Middle Persian in Dari).

I close this section by presenting the vowels of Middle Persian and their basic correspondents in the present time in the three dialects of Persian.

<i>Middle Persian</i>	<i>Dari</i>	<i>Modern Persian</i>	<i>Tajik</i>
i	e	e	i
\bar{i}	\bar{i}	i	i
\bar{e}	\bar{e}	i	e
u	o	o	u
\bar{u}	\bar{u}	u	u
\bar{o}	\bar{o}	u	\bar{u}
a	a	a	a
\bar{a}	\bar{a}	a	o

Table 1: The vowels of Middle Persian and their basic correspondents in the present time in its three main dialects

² Modern Persian has lost final -a, which changed to -e in all words, over time –only two words in Modern Persian end in -a: *va* ‘and’, and *na* ‘no’ (for discussion on neutralization of final -a to -e see Rohany Rahbar, 2007) . In Dari and Tajik, however, final -a is retained from Middle Persian. Compare *kūhna* in Tajik with *kohne* in Modern Persian.

³ These are *del* and *bot* in Modern Persian.

⁴ In Modern Persian, they are *bina* and *bana/bena*.

3. Changing contrasts in the Persian vowel system

In this section, the diachronic changes in the vowel inventory of Persian from the middle to the modern era in terms of contrastive features will be discussed. As a starting point, we must decide on the active features in the vowel system of Middle Persian from which we can track the changes to the inventory of the present time. Two points are important to take into consideration: the feature assignment (that is, which features are active) and also the ordering of them.

It seems that dividing the vowels in the Middle Persian/early Modern Persian inventory into two height classes, low and non-low, can be the first cut in the inventory. This is supported by the following evidence: (i) non-low vowels interact with each other (lowering of *i* and *u* to *e* and *o* respectively, and merging of *ē* and *ō* with *ī* and *ū* respectively) but low vowels do not interact with non-low vowels; (ii) The change that the two low vowels underwent is different from the changes in the vowels in the pair of *i* and *ī* and also in the vowels in the pair of *u* and *ū*. The vowels *i* and *u* underwent a height change to result in *e* and *o*. That is, the distinguishing feature between the two vowels in the original pairs of *i* and *ī* and also *u* and *ū* was length, but due to the lowering of the long vowels, the distinguishing feature between *i* and *e* and also between *u* and *o* became height. In the development of the pair *a* and *ā*, however, length is replaced by place and not by height. After entering [low] as the first feature into the inventory, the result is as follows:

(4)

<i>i</i> <i>ī</i>	<i>u</i> <i>ū</i>
<i>ē</i>	<i>ō</i>
<div style="display: flex; justify-content: space-around; width: 100%;"> <i>a</i> <i>ā</i> low </div>	

Choosing [long] as the second cut seems logical since we consider the inventory to be quantity based. Note that it is also possible to consider [long] to make the first cut, in which case [low] will make the second cut. We will get the same result by applying [low] > [long] or [long] > [low]. The main point is that these two features should be applied before other features. As a result of adding [long] to the system, we will get the following:

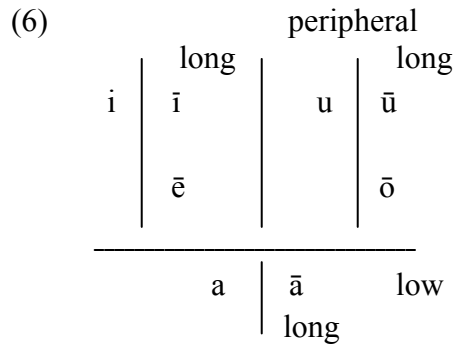
(5)

<i>i</i>	<i>ī</i>	<i>u</i>	<i>ū</i>
	<i>ē</i>		<i>ō</i>
<div style="display: flex; justify-content: space-around; width: 100%;"> <i>a</i> <i>ā</i> low </div>			
	long		

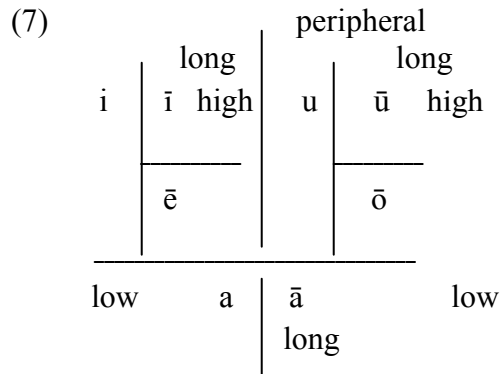
Afterwards, we can apply first [peripheral] and then [high] or the opposite order, that is first [high] and then [peripheral]⁵. Either way will give us the same result. Assume [peripheral] to be

⁵ I use [peripheral], following Rice (1995, 2002), to replace the features [back] and [round] in vowels.

prior to [high]. By adding [peripheral], we can distinguish /i/ from /u/, and /ī, ē/ from /ū, ō/, as follows:



The last cut is made by [high] to distinguish ē from ī and also ō from ū, as follows:



This shows that four features need to be active in the system. The values of these features are given below.

(8)

	a	ā	i	ī	u	ū	ē	ō
low	+	+	-	-	-	-	-	-
long	-	+	-	+	-	+	+	+
peripheral			-	-	+	+	-	+
high				+		+	-	-

It is important to note that in this system *i* and *u* are not contrastively high, thus they could have allophones *i/e* and *u/o*. This means that lowering is redundant to begin with. In addition, note that place is redundant for low vowels. That is, the inventory shows a variation *ā/a* for *a*, and a variation *ā/a* for *ā*.

To sum up, I consider the order [low], [long] > [peripheral], [high] for the vowel system of Middle Persian. The feature [low] is important because all vowels should be distinguished as either low or non-low; [long] is contrastive throughout the system; [peripheral] is, however, limited to non-low vowels, and [high] is only contrastive among long vowels. This system changed in different ways to get to the present systems of Dari, Modern Persian, and Tajik. Dari

and Modern Persian took the same path, with the difference that the merging of long non-low vowels which has been completed in Modern Persian is still on-going in Dari. Tajik, however, took a different path. These three dialects will be discussed below. I start with Dari which is followed by a discussion on Modern Persian. I close this section with Tajik.

3.1. Dari

Consider the Middle Persian inventory which was just discussed (see (7)) and let us see how it changed to result in the present Dari system. We start with low vowels. In the Dari inventory, as given by Pisowicz, the two low vowels are distinguished based on quantity in this dialect. In non-low vowels, the only change that occurred from Middle Persian is lowering of *i* and *u*. The result is the present Dari system, which is given below. The feature values for the Dari system are as previously given in (8) for the Middle Persian vowel system. Given the system in Middle Persian, *i* and *u* do not reflect any contrast to determine whether they should be considered as phonologically high (*i*, *u*) or mid (*e*, *o*). They show an allophonic variation, as mentioned above. Dari is a dialect which shows *e* and *o* allophones.

(9)

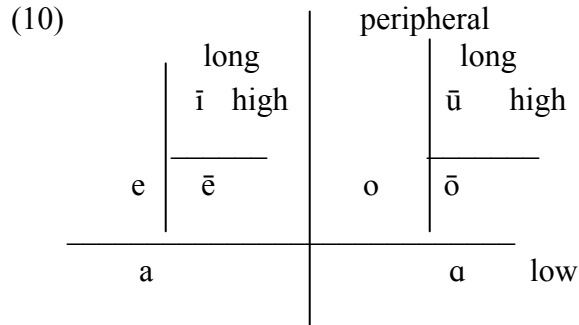
	long ī high		peripheral ū high
e	ē	o	ō
	a	ā	low long

In fact, with the same order as the one of Middle Persian, that is, with [low], [long] > [peripheral], [high], we can account for Dari. No change in contrastive features, therefore, occurred from Middle Persian to Dari. Now I move on to Modern Persian.

3.2. Modern Persian

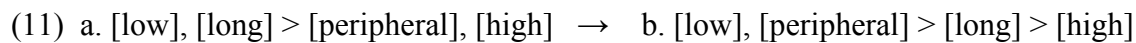
As mentioned above, the inventory of Middle Persian must have undergone three changes to result in the inventory of Modern Persian: (i) fronting of *a* and backing of *ā*; (ii) lowering of *i* and *u*; and (iii) merging of *ē* with *ī*, and merging of *ō* with *ū*. Evidence from Pisowicz (see the reference) seems to show that the changes in low vowels were prior to the other changes, so they had front/back allophones in early Modern Persian. The fronting and backing in low vowels, which happened first, can be taken as a starting point for elimination of quantity. But before elimination of quantity from the system, which happened at a later stage, there is an ambiguity as to whether low vowels are distinguished from each other by [long] or by [peripheral]. I consider [peripheral] to be the distinctive feature for them from this stage and therefore take the low vowels to be *a* and *ɑ*.

Now consider the non-low vowels. If Modern Persian followed the Dari development in terms of order of changes, then we only need to add merging of \bar{e} with \bar{i} , and \bar{o} with \bar{u} . We do not have to assume so, however. Two different dialects do not need to necessarily show the same order of changes (see Dresher, 1990). But the question is: do we have evidence of Modern Persian having taken a different path in term of order of changes? If not, which seems to be the case, we consider the same order of changes for Persian as we observe in Dari. Thus, in Modern Persian, after fronting of a and backing of \bar{a} which result in a and α , we consider lowering of i and u , as follows:



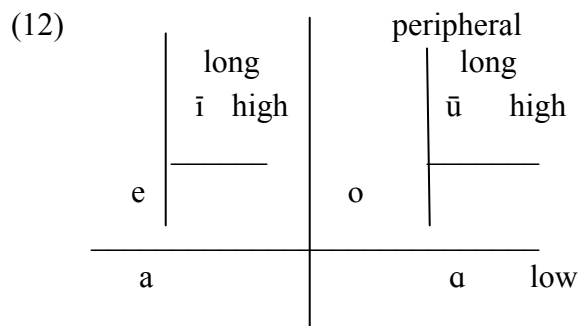
This change does not involve a contrastive feature. The change, resulting in e and o , is, in fact, allophonic for i/e variation and for u/o variation, respectively.

Focusing on the contrastive changes, the first step is reinterpreting the contrast between a and \bar{a} , which is in fact ambiguous, showing a quantity contrast or a place contrast, as a place contrast, as previously mentioned. This means that [peripheral] moves up in the order of features, as shown below, from (11a) to (11b):



Note that in (11b), the order of the features [low] and [peripheral] does not matter. Applying [long] before [high] is, however, crucial in order to keep apart non-low vowels.

The next change in the inventory is merging of the long non-low vowels (i.e., \bar{e} with \bar{i} , and \bar{o} with \bar{u}). The result is the following system:



Now we are faced with an ambiguity as whether the contrastive feature in non-low vowels is quality or quantity. Evidence from Persian synchronic phonology argues for quality to be the

As a starting point, consider the active features of the Middle Persian vowel system with the order [low], [long] > [peripheral], [high]. The first step in Tajik, as in Modern Persian, was to reinterpret the contrast between *a* and *ā* as a place contrast (i.e., [peripheral]) resulting in *a* and *ɑ* for which quantity is redundant. The second step is reinterpreting *i/ī/ē* and *u/ū/ō* in the following way: in Tajik, the cut made by [high] was prior to the cut made by [long]; that is, [high] > [long]. Recall the allophonic variation *i/e* and *u/o* that we discussed for the Middle Persian vowel system. Tajik was a dialect which showed *i* and *u* allophones. The former Tajik system can be, therefore, represented as follows:

(17)

i	long ī high	u	peripheral long ū high
	ē		ō
	a		ɑ low

In this system, as observed above, *ē* and *ō* are not contrastively long. When the length distinction is lost in Tajik, it results in the merger of *i* with *ī* and also *u* with *ū*. The result is a vowel system like the system of Modern Persian, given above in (14), repeated here as (18).

(18)

i		peripheral u high
e		o
a		ɑ low

The inventory in (18) underwent two further changes to result in the Modern Tajik inventory, which was given above in Figure 1 and is repeated here in (19). These two changes are as follows: (i) *ɑ* → *o*, and (ii) *o* → *ũ*.

(19) The Tajik vowel inventory

i	u
e	ũ o
a	

The question is: how did the two changes, *ɑ* → *o*, and *o* → *ũ*, occur in the inventory of Tajik? Recall that the Tajik vowel system is believed to have been influenced by Turkic. Studies show that among the Turkic languages, Uzbek has been to a great extent in contact with Tajik (e.g., Sjoberg, 1963; Boeschoten, 1998; Johanson, 1998; Bodrogligeti, 2003).

Considering the influence of Uzbek on Tajik, it is worthwhile to try a Turkic-style inventory for Tajik, to examine whether the presence of *ı̄* and rounding of *a* to *o* can be accounted for with such an inventory. It should be noted that although it is repeatedly mentioned in literature that Tajik has been under influence of Turkic, in particular Uzbek, the nature of this influence has not been explained.

Before trying a Turkic-style inventory for Tajik, let us take a look at the Uzbek vowel system. We are, in particular, looking for some explanation for Tajik *ı̄* and rounding of *a* to *o*.

According to Harrison, Dras, and Kapicioglu (2002) the old Uzbek vowel system was as follows:

(20) The Old Uzbek vowel inventory⁶

i	y	ɯ	u
e	ø	a	o

They afterward mention that *y, ø, ɯ* underwent merger and disappeared and therefore Modern Uzbek is left with five vowels. According to Sjoberg, in Uzbek, *ı̄* is a high back-central, rather close, slightly rounded allophone of /u/. Thus the inventory of Modern Uzbek can be represented as follows:

(21) The Modern Uzbek vowel inventory

i	(ı̄)	u
e	a	o

Considering the Uzbek vowel system, we can speculate that Tajik speakers reinterpreted the Persian vowel system in Turkic terms. That is, a change of a three-height system to a two-height system occurred, in the following way: (i) *e* is considered phonologically low; (ii) *a* moves towards the center; (iii) *a* goes to *ɔ*, which is low *o*. Now *o* has to maintain contrast with *a* which merged with *ɔ* and with *u*. The contrast cannot be maintained by height. This leads to the last change, as follows: (iv) *o* occupies the empty space resulting in *ı̄*. As for the rounding of *a* to *o*, Rona-Tas (1998) and Johanson consider the letter <o> in Uzbek to represent a labialized or back /â/. A similarity between Uzbek and Tajik is, thus, observed with regard to the low back vowel being rounded. I do not go further through Uzbek since to identify contrastive features in the Uzbek vowel system is not a goal in this paper. The important point for our study was to find an explanation for the changes in the Tajik vowel system considering the impact of Uzbek on Tajik, which we just discussed.

⁶ In the reference, the vowels are given in two columns, which indicate front and back distinction. Most likely the reason to choose columns is that the authors indicate the frequency of each vowel in front of it. I put them in two rows as for typical Turkic inventories.

In a two-height Turkic-style inventory, Tajik does not need to have two distinctive height features. In a Turkic-style vowel inventory, then, we can account for the Tajik vowel system with the order: [low] > [peripheral], [coronal], as follows:

(22)

coronal		peripheral	
i	(ũ)	u	
e	a	o	low

In this inventory, the feature values are as follows:

(23)

	a	e	i	ũ	o	u
low	+	+	-	-	+	-
peripheral	-	-	-	-	+	+
coronal	-	+	+	-		

This inventory shows two distinctive place features (i.e., [peripheral] and [coronal]) and one height feature (i.e., [low]).

More work is needed to confirm the Turkic-style inventory that we speculated for Tajik.

4. Conclusion

I provided an analysis based on changing contrasts in the vowel systems of three main dialects of Persian: Dari, Modern Persian, and Tajik taking Middle Persian as the starting point.

I showed that no contrastive change occurred from Middle Persian to Dari. The quantitative vowel system of Middle Persian was, however, replaced by a qualitative system in Modern Persian and Tajik. In addition, Tajik, under the influence of Turkic, has its own particular characteristics to account for which a two-height Turkic-style inventory was suggested.

The changes of contrasts from Middle Persian which results in the vowel systems of the three dialects of Persian are as follows:

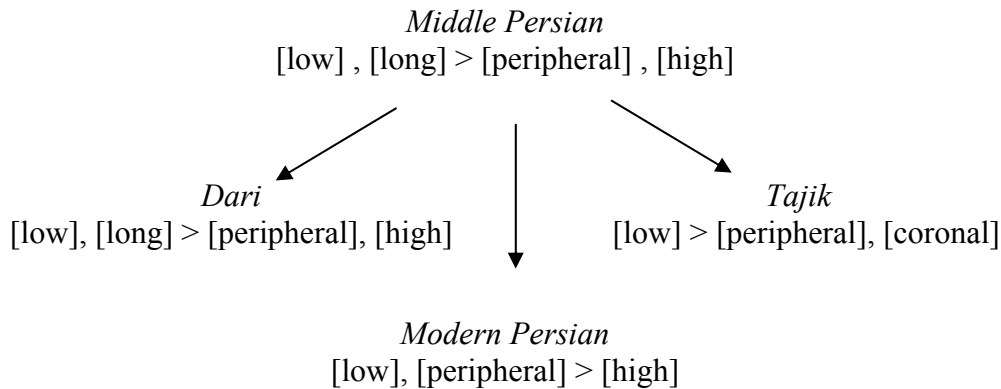


Figure 2: The development of the Middle Persian vowel system to its current main dialects in terms of contrastive features

References

- Bahar, Muhammad Taqi (1942). *Sabk Shenasi*. Tehran: Khodkar Publishers.
- Bodrogligeti, Andras J. E. (2003). *An Academic Reference Grammar of Modern Literary Uzbek*. LINCOM Europa.
- Boeschoten, Hendrik (1998). Uzbek. *The Turkic Languages*, eds. Lars Johanson and Eva Agnes Csato. MPG Books Ltd, Bodmin, Great Britain.
- Dresher, B. Elan (1990). On the Unity of the Mercian Second Fronting, in Sylvia Adamson et al. (eds.) *Papers from the 5th International Conference on English Historical Linguistics (Current Issues in Linguistic Theory 65)*, John Benjamins, Amsterdam, 141-164.
- Dresher, B. Elan (2003a). Contrast and Asymmetries in Inventories. In Anna-Maria di Sciullo, ed., *Asymmetry in Grammar, Volume 2: Morphology, Phonology, Acquisition*, John Benjamins, Amsterdam, 239-257.
- Dresher, B. Elan (2003b). Determining Contrastiveness: A Missing Chapter in the History of Phonology. In Sophie Burelle and Stanca Somesfalean, eds., *Proceedings of the CLA*.
- Dresher, B. Elan (2003c). The Contrastive Hierarchy in Phonology. In Daniel Currie Hall, ed., *Toronto Working Papers in Linguistics (Special Issue on Contrast in Phonology)* 20:
- Dresher, B. Elan, Glyne Piggott, and Keren Rice (1994). Contrast in phonology: Overview. *Toronto working Papers in Linguistics* 13: iii-xvii.
- Dresher, B. Elan and Xi Zhang (2005). Contrast and phonological activity in Manchu Vowel Systems. *Canadian Journal of Linguistics* 50 (1-4).
- Harrison, David K., Mark Dras, and Berk Kapicioglu (2002). Agent-based modeling of the evolution of vowel harmony. *Proceedings of NELS* 32: 217-236. New York.
- Johanson, Lars (1998). The history of Turkic. *The Turkic Languages*, eds. Lars Johanson and Eva Agnes Csato. MPG Books Ltd, Bodmin, Great Britain.
- Kalbasi, Iran (1995). *Farsi-y-e Iran and Tajikestan*. Tehran: Daftar-e motaleat-e Siyasi va Beynolmelali.
- Meshkatod Dini, Mehdi (2000). *Tosif and Amouzesh-e Zaban-e Farsi*. Ferdowsi University Press. Mashad, Iran.
- Natel Khanlari, Parviz (1987). *Tarikh-e Zaban-e Farsi*. Tehran: Nashr-e No.
- Perry, John R. (2005). *A Tajik Persian Reference Grammar*. The Netherlands: Koninklijke Brill NV.
- Pisowicz, Andrzej (1985). *Origins of the New and Middle Persian Phonological system*. Nakladem Uniwersytetu Jagiellonskiego.
- Rastorgueva, Vera Sergeevna (1969). *Dastur-e Zaban-e Farsi-y-e Miyane*. Tehran: Bonyad-e Farhang-e Iran.
- Rice, Keren (1995). On vowel place features. *Toronto working papers in Linguistics* 14: 73-116.
- Rice, Keren (2002). Vowel place contrasts. In M. Amberber and P. Collins (eds.), *Language universals and variation*. Westport, Connecticut: Praeger. 239-69.
- Rohany Rahbar, Elham (2007). Vowel Height in Persian. Ms. University of Toronto.
- Rona-Tas, Andras (1998). Turkic writing systems. *The Turkic Languages*, eds. Lars Johanson and Eva Agnes Csato. MPG Books Ltd, Bodmin, Great Britain.
- Salemman, Karl Germanovich (1930). *A Middle Persian Grammar*. Mazgaon: British India Press.

- Samareh, Yadollah (1977). *The arrangement of segmental phonemes in Farsi*. University of Tehran Press.
- Samareh, Yadollah (1985). *Avashenasiye Zabane Farsi*. Markaze nashre daneshgahi. Tehran, Iran.
- Sjoberg, Andree F (1963). *Uzbek Structural Grammar*. Mouton and Co., The Hague. The Netherlands.
- Windfuhr, Gernot L. (1979). *Persian Grammar: History and State of its Study*. Trends in Linguistics 12. Mouton Publishers.

Author contact information:

Elham Rohany Rahbar: elham.rohanyrahbar@utoronto.ca