

Azeri Morphosyntactic Variation: The Effect of Persian on NP Structures

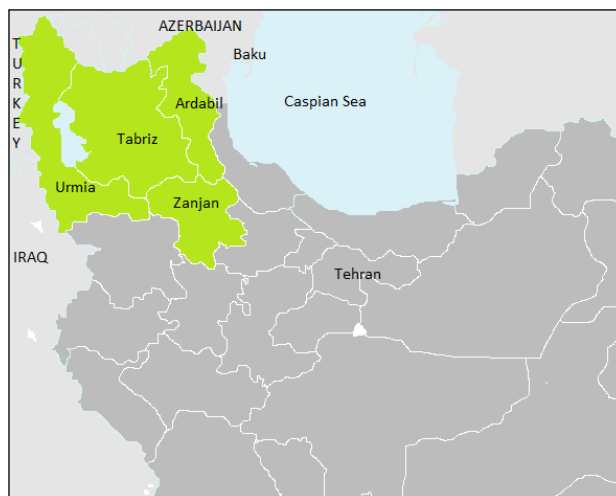
Parisa Erfani
Simon Fraser University

1. Introduction

Iran is a diverse country, with people of many religious and ethnic backgrounds who speak different languages as their first language. Persian is the dominant language and native speakers of Persian often do not learn a minority language. However, most speakers of minority languages in Iran also speak Persian to some degree. Persian serves as a *lingua franca* in Iran, and most publications and mass media are in this language. There is only limited publication or broadcasting programs in the other relatively popular languages of Iran, such as Azeri and Kurdish. In some societies, people use one language in their families, local communities, and work, but another language for education and official business. This is the situation in Iran: the only official language of Iran is Persian, and it is the only language used for education, including in Azeri-speaking areas. Many educated Azeris are totally fluent in both Azeri and Persian. Equally comfortable in both languages, bilingual speakers often engage in code-mixing when speaking to each other.

Azeri is classified as a Turkic language. There are two main varieties of the language can be recognized based on different cultural and linguistic regions. One variety, which is spoken in the Republic of Azerbaijan, is called Azerbaijani and the other variety, which is spoken in Iran, is called Iranian Azeri or Azeri. This study focuses on the Azeri language as it is spoken in Iran. Azeri is a Turkic language, but it is strongly influenced by Persian, an Indo-European language. Azeri, with approximately 15–20 million speakers, has more speakers than any other non-Persian language in Iran (Crystal 2010). Most Azeri speakers inhabit the four provinces in the northwestern part of Iran. Each province has its own dialect—the Ardabil dialect in Ardabil province, the Tabriz dialect in East Azerbaijan province, the Urmia dialect in West Azerbaijan province and the Zanzan dialect in Zanzan province. The dialects are mutually intelligible, although they are distinguished by phonological and lexical criteria (Dehghani 2000). I am a native speaker of Azeri, born and raised in Tabriz, capital of East Azerbaijan province in northwestern Iran.

Image 1. Map of Azeri-speaking areas in Iran¹



Given the dominance of Persian, and the long period of intensive contact, Azeri speakers differ in their fluency in Persian, ranging from monolinguals to fully functional bilinguals. People from the older generation who have little or no education are not able to read, write or speak Persian fluently. However, those who have higher education, which includes most of the younger generation, can read, write and speak Persian fluently. The reason is that they have been in contact with Persian for many years, they read academic publications in Persian, and of course, many of the educated people need to write academic texts.

Borrowing is not limited to lexical items. Myers-Scotton (1993) states that when two languages that are not genetically related share a geographical location, and there is a high degree of bilingualism or multilingualism, grammatical features of the dominant language may be adopted by the minority language. Since, Persian is the only official language in Iran it has political and cultural dominance over Azeri. This is exactly the sort of situation where one would expect the structure of a language to be influenced by another language, even if it is typologically dissimilar. Erfani (2012) explored this issue for a variety of morphosyntactic constructions in Azeri and found that several show signs of persification.

The main objective of this paper is to examine patterns of language variation among Azeri speakers in their use of relative clauses and compound nouns in order to determine the degree of influence of Persian on Azeri structure. To do this, I designed a study to investigate Azeri relative clause and compound noun constructions, collecting data from a variety of Azeri speakers. Section 2 gives an introduction to relative clauses in Azeri followed by an introduction to noun compounding in Azeri in section 3, as compared to Turkish and Persian. Section 4 describes the field study detailing the methodology. Section 5 analyses the relative clause and compound noun data and discusses the results in terms of two sociolinguistic factors—the age and level of education of the speaker. Finally, section 6 summarizes the results of this study and discusses what it reveals for the

¹ This map, retrieved November 1, 2012, is constructed using the map template from <http://en.wikipedia.org/wiki/File:Blank-Map-Iran-With-Water-Bodies.PNG>

future of the Azeri language.

2. Relative Clauses

Typological studies investigating relative clauses have established that there is a strong correlation between the basic word order in a language and the position of the relative clause in relation to the head noun (Downing, 1978; Keenan 1985). Relative clauses are categorized typologically with respect to the position of the head noun; pre-nominal RCs precede the head and post-nominal RCs following the head.

There are several ways of forming pre-nominal RCs two most common ways involve the relativizer suffixes *-(y)An* or *-dIK*, which are suffixed to a non-finite verb:²

- (2) [(män-Ø) sän-ä Øi göstär-an] äksi
 [(I-nom) you-dat show-rel] picture
 'the picture that i showed you'
- (3) [(män-(im)) sän-ä Øi göstär-dığ-im] äksi
 [I-(gen) you-dat show-rel-poss.1sg] picture
 'the picture that I showed you'

Pre-nominal RCs are typical in Turkic languages. As in Azeri, the two most common ways involve the relativizer suffixes *-(y)An* or *-dIK* (Kornfilt 1997; Göksel and Kerslake 2005; Göksel 2009).

- (4) [sirt-in-da çanta ol-an] kız
 [back-poss.3sg-loc bag be-rel] girl
 'the girl who has a bag on her back'
- (5) [Ankara-da gör-düğ-üm] bayan
 [Ankara-loc see-rel-poss.1sg] lady
 'the lady whom I saw in Ankara'

In addition to pre-nominal RCs, Azeri also has post-nominal RCs, presumably arising due to the influence of Persian. In this RC structure, the RC is introduced by the complementizer *ki* 'that' which is a borrowed from Persian *ke* 'that', and connects the head noun to the relative clause. In this structure, which is simpler than the pre-nominal RC structure, the relative clause takes a finite verb without any inflection for case or verbal agreement.

- (6) o oğlani [ki Øi qısa şalvar gey-ib-dir]
 that boy [comp short pants wear-pf.3sg]
 'the boy who is wearing the short pants' (Participant 6: 2012)
- (7) o adam-lar [ki türki kitab-lar-ı oxu-yar-lar]
 that person-pl [comp Turkish book-pl-acc read-aor-pl]
 'those people who read Turkish books' (Participant 4: 2012)

² The relativizers *-(y)An* and *-dIK* are the most common relativizers. There are two more relativizers *-mİş* and *-(y)AcAk*, which are used rarely in Azeri.

Persian has head-initial typology. Persian relative clauses are introduced by the complementizer *ke* ‘that’ and follow the head noun:

- (8) ân doxtar [ke gol dâr-ad]
 that girl [comp flower has-pres.3sg]
 ‘the girl who has a flower’

3. Noun Compounding

Compounding, which is probably the most common morphological process cross-linguistically, can be defined as a lexical item consisting of two or more words used for generic rather than referential function, e.g. English garbage man or popcorn (Fabb 1998: 66). Azeri compound nouns come in two forms: one can be regarded as the native Turkic variant and the other variant is borrowed from Persian. Thus noun compounding can serve as a measure of Persian influence on Azeri. Native Azeri has right-headed noun-noun and adjective-noun compounding:

- (9) mârmâr daş
 marble stone
 ‘marble stone’ (Participant 3: 2012)

- (10) böyük -maman
 big mother
 ‘grandmother’ (Participant 10: 2012)

The above compounds are bare, but for noun-noun compounds, it is more common to use the linker *-(s)I*.

- (11) İsfahlan känd -i
 İsfahlan village -lnk
 ‘İsfahlan village’ (Participant 2: 2012)

- (12) lobya kükü -si
 bean omelet -lnk
 ‘green bean omelet’ (Participant 5: 2012)

Right-headed compound structures are typical in Turkic languages. As in Azeri, the most productive and frequently used compounds in Turkish are noun-noun and adjective-noun (Kornfilt 1997; Göksel and Kerslake 2005; Göksel 2009; Ralli and Bağrıaçık 2011; among others).

- (13) ipek çorap
 silk sock
 ‘silk sock’

- (14) böyük -anne
 big -mother
 ‘grandmother’

Noun-noun compounding can also be formed with an *-(s)I* suffix, as in:

- (15) para çanta -sı
 money bag -lnk
 ‘purse’
- (16) İngiliz edebiyat -ı
 English literature -lnk
 ‘English literature’

Persian also has bare noun-noun and noun-adjective compounds:

- (17) âb -havij
 water -carrot
 ‘carrot juice’
- (18) pedar -bozorg
 father -big
 ‘grandfather’

The above examples are left-headed, which is considered the default order of compounds in Persian (Kalbasi 1992; Shariat 2005; Anvari and Ahmadi-Givi 2006; Mahoozi 2006; Vahidian-Kamyar and Omrani 2006; Foroodi-Nejad and Paradis 2009), though right-headed compounds also occur.

- (19) noxost -vazir
 first -minister
 ‘prime minister’

Another way of forming compounds in Persian is by means of the Ezafe construction.³ The head noun is suffixed with the Ezafe *-(y)e* (the glide *-y-* occurs after vowels).

- (20) daryâ -ye xazar
 sea -ez Caspian
 ‘Caspian sea’
- (21) miz -e utu
 table -ez iron
 ‘ironing table’

Such compounds are left-headed. Persian is a language that has variable head positions in noun compound structures. Azeri speakers also frequently use the left-headed Ezafe construction:

- (22) müdir -i mədrəsə
 director -ez school
 ‘the school director’

(Participant 3: 2012)

³ In Persian, the Ezafe construction with a vowel *-e* occurs with various kinds of post-nominal modifiers, including APs, descriptive NPs, genitive NPs, and some PPs (Samiian 1994).

- (Participant 6: 2012)

These are formed with the Ezafe suffix, which is borrowed from Persian. The above phrases, which are direct quotation from Persian, could alternatively be expressed in Azeri by right-headed equivalents:

- (25) türki dil -i
 Turkish language -lnk
 'Turkish language'

I consider the right-headed compound in Azeri to be the native Turkic pattern since Turkish generally lacks left-headed compounds.

4. Methodology

In order to investigate the morphosyntax of Azeri and the influence that Persian has on it, I travelled to Tabriz, Iran, to conduct a field study. This project is a qualitative/quantitative study designed to compare Azeri as spoken by the younger and older generations. This field research involved ten participants divided into two groups. The participants in the older generation (aged 65+) were mostly monolingual in Azeri and the participants in the younger generation (aged 20–35) were mostly bilingual in Azeri and Persian. They can be further sub-divided by their level of education (basic education or higher education). The interviews were recorded with a high quality digital voice recorder (Olympus WS 801). The participants were each involved in a 30-45 minute free conversation in an informal setting in a quiet room at the participant's home. The interviews resulted in a total of 6 hours and 50 minutes of speech (189 minutes by older speakers, 221 minutes by younger speakers). Selected data were transcribed and translated and these formed the basis of my dataset.

5. Data Analysis

Over the last forty years, language variation theorists have developed a methodology for applying sociolinguistic analysis to the variation found in the phonological, morphological, syntactic and semantic structure of a language. Labov (1972c) defines a linguistic variable as simply “two ways of saying the same thing.” Tagliamonte (2006: 70) refines this notion, saying that the variants should not result from performance anomalies, but be linguistically well-formed. Furthermore, the frequency of variation should be robust: both variants must occur with sufficient frequency. A variationist approach to linguistic analysis can then look for factors that elucidate the systematic distribution of the variants. Ferguson (1959), Calteaux (1994), Thomason and Kaufman (1998) and Thomason (2003) are among those to discuss the effect of social factors in language contact. When speakers of different languages live in close contact, their languages

influence each other, but they do so in piece-meal fashion, leading to complexities in the synchronic language structure and differences among speakers. Variations that gain popularity can gradually lead to loss of a variant and result in language change. According to Labov (1994, 2001), some of the socio-cultural factors that can affect the use of linguistic variables are age, sex, social class, ethnicity, race and community size.

My research seeks to examine language change in progress in the Azeri language by comparing the data from monolingual Azeri speakers to the data from bilingual Azeri-Persian speakers. This study shows that two socio-cultural factors, age and level of education, are relevant to morphosyntactic variation in Azeri. First, we look at the effect of the age and next the effect of education. The age of the speaker has been demonstrated to be an important social factor in language variation (Labov 2000). Differences between generations in linguistic behavior illustrate clear examples of language change in progress. Thus, the age of the speaker becomes an important factor when investigating the status of a linguistic structure in a community. One goal of my field study was to see whether the factor of age influences the choice of compound noun variant.

As stated earlier, Azeri has two relative clause and compound noun variants. My data yielded 85 tokens of RCs: 38 (45%) were pre-nominal and 47 (55%) were post-nominal. In other words, the persified head-initial construction was slightly preferred over the native Turkic head-final construction. In this study, which yielded 225 tokens of CNs, right-headed and left-headed compounds are both robustly attested, with a slight preference for the latter: 43% were right-headed (96 CNs) and 57% were left-headed (129 CNs). In other words, the persified left-headed CNs was slightly preferred over the native Turkic right-headed construction.

These results suggest that relative clauses and compound nouns provide good linguistic variables to investigate because both variants are produced frequently in daily speech. Given the results of the RC and CN data above, an obvious question to ask is whether the social factors of age and education influence the choice of variants in relative clause and noun compounding.

5.1 *Effect of Age*

The following tables give break-down in the results of the two types of the relative clauses and compound nouns as produced by older and younger groups of speakers.

Table 26. Number and percentage of pre-nominal and post-nominal RCs by older and younger groups

Participants	pre-nominal		post-nominal		Total
	#	%	#	%	#
older group	31	68	15	42	46
younger group	7	18	32	82	39
Total	38	45	47	55	85

As seen in table 26, the older speakers produced 31/46 pre-nominal RCs and 15/46 post-nominal RCs, whereas the younger speakers produced 7/39 pre-nominal RCs and 32/39 post-nominal RCs. Therefore, the results show that older participants tend to produce more pre-nominal RCs (68%), whereas the younger participants tend to produce

more of the post-nominal variant (82%). In sum, the total number of 38/85 of the pre-nominal variant (45%) and 47/85 of the post-nominal variant (55%) were found and transcribed.

Table 27. Number and percentage of right-headed and left-headed compound nouns by older and younger groups

Participants	right-headed		left-headed		Total
	#	%	#	%	#
older group	51	58	37	42	88
younger group	45	33	92	67	137
Total	96	43	129	57	225

As table 27 illustrates, the older speakers produced 51/88 right-headed compounds and 37/88 left-headed compounds, whereas the younger speakers produced 45/137 right-headed compounds and 92/137 left-headed compounds. Therefore, the results show that older participants tend to produce more of the right-headed compound noun variant (58%), whereas the younger participants tend to produce more of the left-headed borrowed variant (67%). The older participants tend to produce slightly more compounds with native Azeri structures than with the borrowed Persian order, whereas the younger participants tend to produce more compounds with the borrowed structure than with the native one.

5.2. Effect of Education

In the sociolinguistic literature, many studies have been done on the effect of education on language variation. Education may be the best factor measuring the social evaluation of features in a community, with higher levels of education correlating with linguistic features held to have prestige (Labov 2002: 60). In this study, the effect of education has been investigated differentiating between participants with little or no education versus those with some post-secondary education. The following tables present the number and percentage of pre-nominal versus post-nominal RCs and right-headed versus left-headed CNs tabulated for two groups of speakers—those with little or no education and those with higher education.

Table 28. Number and percentage of pre-nominal versus post-nominal RCs by level of education

Participants	pre-nominal		post-nominal		Total
	#	%	#	%	#
less educated	28	76	9	24	37
higher educated	10	21	38	79	48
Total	38	45	47	55	85

As table 28 shows, the less educated speakers produced 28/37 post-nominal RCs and 9/37 pre-nominal RCs, whereas the more educated speakers produced 10/48 pre-nominal RCs and 38/48 post-nominal RCs. That is, less educated participants tended to produce more of the head-final variant (76%) whereas the more educated participants tended to produce more of the head-initial variant (79%). These statistics show that the less educated

participants favor the native Azeri structure. In contrast, the behaviour of educated speakers is the opposite—they tend to produce RCs with the borrowed structure.

Table 29. Number and percentage of right-headed and left-headed compound nouns by level of education

Participants	right-headed		left-headed		Total
	#	%	#	%	#
less educated	40	71	16	29	56
higher educated	56	33	113	67	169
Total	96	43	129	57	225

Table 29 illustrates, the less educated speakers produced 40/56 right-headed compounds and 16/56 left-headed compounds, whereas the more highly educated speakers have produced 56/169 right-headed compounds and 113/169 left-headed compounds. The results show that the less educated participants tend to produce more of the right-headed variant (71%), whereas the more highly educated participants tend to produce more of the left-headed variant (67%). This statistic shows that the less educated participants favor the native Azeri structure. In contrast, the behavior of the educated speakers shows that they tend to produce more compounds with the borrowed structure.

To summarize, the findings in the present study show that noun compounds are a good sociolinguistic variable in Azeri because both right-headed and left-headed compound nouns are well attested. The data show that the factors of age and education influence the choice between variants. Summarizing the results overall, young and educated speakers, who have more contact with Persian through media, education and social contact, are more influenced by Persian structure. In contrast, older speakers, who are mostly monolingual and have less education in the Persian language, retain more native Azeri structures in their speech.

6. Conclusion

This study examines linguistic issues in Azeri, the effect of Persian on Azeri morphosyntax. Iranian Azeri has been strongly influenced by Persian, an Indo-European language. Intensive linguistic and cultural contact has led to considerable convergence between the two languages. Northwestern Iran is an ethno-linguistic contact zone where Azeri and Persian have been spoken side by side for more than a millennium.

We saw that in noun compounding, left-headed and right-headed compound nouns were used with almost equal frequency by the participants. However, the choice of structure differed slightly by the age and education of the participants. The finding of the current study is compatible with the findings of other studies on languages of the region. Johanson (1998) claims that persification in the Irano-Turkic area is promoted by increased education and communication. These findings also show that Azeri is becoming persified, as predicted in situations of language contact involving a politically-dominant language. An interesting future study would be to compare the status of Azeri to other varieties of the Azerbaijan language, particularly Northern Azerbaijani, the official language in the Republic of Azerbaijan.

According to the results of my study, the influence of Persian is seen to be greater among young, educated speakers. With respect to the factor of age, Sankoff and Thibault (1981) claim that if a syntactic variant is correlated with age, this may be evidence of language change in progress. For example, left-headed variant correlates with the younger group and thus this might be an indication of an evolution in the grammar of Azeri toward Persian structure. Sankoff and Thibault (1981) further argue that when variants coexist for a long time, it should be expected that this equivalence will be grammaticalized at a later time. Therefore, we should expect structures such as left-headed compound nouns, which has been borrowed from Persian and has coexisted with native Turkish structure for a long time, will be eventually be considered as canonical structures in the grammar of Azeri.

Furthermore, the difference between the two groups of speakers in my study suggests that the rate of persification of Azeri is accelerating. However, due to the small number of participants and tokens, these conclusions can only be suggestive. Additional quantitative studies with sufficient data are required to verify these results. This discovery is an issue of some concern. The topic of language endangerment often focuses on languages with small populations of people, e.g. indigenous languages of North America. But even when a language is spoken by millions of people, it can undergo rapid decline in the face of contact.

References

- Anvari, H. & Ahmadi-Givi, H. (2008). *Persian Language Grammar 2*. Tehran: University of Tehran Publications.
- Calteaux, K. (1994). *A sociolinguistic analysis of multilingual community*. Ph.D. dissertation, Rand Afrikaans University.
- Crystal, D. (2010). *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.
- Dehghani, Y. (2000). *A grammar of Iranian Azeri: including comparisons with Persian*. Munich: Lincom Europa.
- Downing, B. (1978). Some universals of relative clause structure. In Greenberg, (Ed.), pp. 376-418.
- Erfani, P. (2012). Azeri morphosyntax: The influence of Persian on a Turkic language. MA thesis, Simon Fraser University.
- Fabb, N. (1998). Compounding. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphology*, pp. 66-83. Oxford: Blackwell.
- Ferguson, Ch. A. (1959). Diglossia. *Word*, 15, 325-340.
- Foroodi-Nejad, F. & Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in Persian-English bilinguals. *Bilingualism: Language and Cognition*, 12(4), 411-427.
- Göksel, A. & Kerslake, C. (2005). *Turkish, a comprehensive grammar*. London: Routledge.
- Göksel, A. (2009). Compounds in Turkish. *Lingue e Linguaggio*, 2, 213-236.
- Johanson, L. (1998). Iranian elements in Azeri Turkish. In E. Yarshater (Ed.), *Encyclopaedia Iranica* 3, pp. 248b-251a. London, New York.
- Kalbasi, I. (1992). *Saxt-e eshteghagh-yē vazhe dār Farsi-yē emruz* [Word derivational structure in Modern Persian]. Tehran: Mo'assesse-ye Motale'at va Tahghighat-e Farhangi (Pazhuheshgah).

- Keenan, E. L. (1985). Relative clauses. In T. Shopen (Ed.), *Language Typology and Syntactic Description. 2: Complex Constructions*, pp. 141-170. Cambridge: Cambridge University Press.
- Kıral, F. (2001). *Das gesprochene Aserbaidshisch von Iran: Eine Studie zu den syntaktischen Einflüssen des Persischen*. Wiesbaden: Harrassowitz Vorlage.
- Kornfilt, J. (1997). *Turkish*. London: Routledge.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1994). *Principles of linguistic change. Volume I: Internal factors*. Oxford: Basil Blackwell.
- Labov, W. (2000). The anatomy of style shifting. Stylistic variation in language. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation*, pp. 85-108. Cambridge: Cambridge University Press.
- Labov, W. (2001). *Principles of linguistic change. Volume 2: social factors*. Malden, MA: Blackwell.
- Labov, W. (2002). Driving forces in linguistic change. Paper presented at the 2002 *International Conference on Korean Linguistics*. Seoul National University. Seoul, South Korea.
- Lee, S. N. (2008). *A grammar of Iranian Azerbaijani*. Seoul: Thaeaksa.
- Mahoozi, M. (2006). *Ketab Farsi omoomi* [Persian grammar]. Tehran: Zavar Press.
- Myers-Scotton, C. (1993). *Dueling languages: grammatical structure in codeswitching*. Oxford: Clarendon.
- Ralli, A. & Bağrıaçık, M. (2011). Bare N(ominal) N(ominal) concatenations in Turkish: Compounds or syntactic falacies? In N. Hathout & F. Montermini (Eds.), *Morphologie à Toulouse*. München.
- Samiiian, V. (1994). The Ezafe Construction. In M. Marashi (Ed.), *Persian Studies in North America*. Bethesda, Maryland: Iranbook.
- Sankoff, D & Thibault, P. (1981). Weak complementarity: tense and aspect in Montreal French. In B. B. Johns & D. R. Strong (Eds.), *Syntactic change. Natural Language Studies* 25, pp. 205-216.
- Tagliamonte, S. A. (2006). *Analyzing sociolinguistic variation*. London: Cambridge University Press.
- Thomason, S. G. & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Thomason, S. G. (2003). Contact as a source of language change. In R. D. Janda & B. D. Joseph (Eds.), *A handbook of historical linguistics*, pp. 687- 712. Oxford: Blackwell.
- Vahidian, K. & Omrani, Gh. (2006). *Dastoor zaban Farsi (1)* [Persian grammar]. Tehran: Samt Press.