

Relative clauses in Persian

A small-scale corpus study

Elias Abdollahnejad and Hamideh Marefat

University of Calgary / University of Tehran

Introduction

Relative clauses (RC) as one of the complex subordinate structures of language have been the focus of extensive study. The present study is a corpus based investigation of Persian RCs. The main aim is to investigate the distribution of different RC types and their typical syntactic characteristics in the present-day formal written Persian. The paper is organized as follows: first we introduce Persian RCs and the different types they have. Then, the claims regarding the position of the direct object marker *râ*, whether it comes after the complex DP or between the head noun and the RC, and the frequency of the occurrence of the main verb between the head DP and its modifying clause are examined based on the corpus. Next, Comrie's (1989) suggestion on various realizations of the relativized element (repetition, pronoun retention, topicalized pronoun, and gap) inside the modifying clause is used to classify the RCs. Later, the effect of RC type and relativized element on the distribution of gaps, resumptive pronouns (RP henceforth), and pronominal clitics suggested by Taghvaipour (2005) is scrutinized. Also, his proposal on the behavior of coordinated RCs regarding the occurrence of gaps or RPs inside each conjunct is taken into account. To sum up, this study will be an attempt to answer the following research questions:

1. Do Persian speakers prefer to put the direct object marker *râ* between the head noun and the RC or after the complex DP?
2. Do Persian speakers prefer to put the main verb between the head DP and its modifying clause or at the end of the complex DP?
3. What are the frequencies of different possible realizations of the relativized element (repetition, pronoun retention, topicalized pronoun, and gap) inside the modifying clause in Persian RCs?

4. What is the frequency of gaps, RPs, and clitics across different types of Persian RCs?
5. What is the behavior of coordinated Persian RCs regarding the occurrence of gaps or RPs inside each conjunct?

1. Review of the literature

1.1 Relative clauses in Persian

Persian RCs, like English ones, are DP initial (i.e. the modified DP precedes the modifying clause). An example of a Persian Subject RC is provided below (1). In the Persian examples, OM stands for the object marker particle *râ* always following [+specific] direct objects (Karimi, 2001) which can be considered as an enclitic (Dabir-Moghaddam, 1992) and we use “=” to show its attachment to its prosodic host;¹ *ps* and *pr* refer to past tense and present tense, respectively. Moreover, *DEM* stands for demonstrative *-i* sometimes connected to the head DP.

- (1) *ketâb-i [ke u pišnæhâd kærd]=râ xærid-æm.*
 book-DEM ke (s)he proposal do-ps-3sg=OM buy-ps-1sg
 (I bought the book that she proposed.)

Particle *ke* is obligatorily used with most of the Persian RCs to link the RC to the matrix sentence (Windfuhr, 1979). Karimi (2001, p. 11) considers this *ke* particle as an “invariant relative complementizer” introducing Persian relative clauses. Pointing to the necessity of this particle in relative clauses and providing a short list (which he believes is incomplete) of its properties, Darzi (2008, p. 111) claims that this particle, which introduces Persian embedded clauses, “behaves more or less like a complementizer”.

In this paper we use the RC classification based on the syntactic function of the relativized element inside the RC. These relativized elements can have the syntactic function of subject, object, oblique, possessive, or adjunct (of time or place) in the RC structure. Thus these RCs are named after the syntactic function of their relativized element.

1.2 Two major types of RCs

Quite like English, there are two major types of RCs in Persian: ordinary RCs (restrictive or nonrestrictive) modifying a head noun in the matrix clause, as in

1. We should thank an anonymous reviewer for bringing this point to our attention.

(1) above, and free RCs, as in (2) below, which rather than modifying a head noun stand alone and do not possess an overt head. Taghvaipour (2005, p. 193) proposes that Persian free RCs always begin with “the prefix *hær-* ‘-ever’ ” linked to a *wh*-word. Sentence (2) illustrates the ungrammaticality of *wh*-words (e.g. *ki* ‘who’) without prefix *hær-* in Persian free RCs.

- (2) [*hærki*/**ki* *u=râ* *did*] *sælâm kærd*.
 whoever/who (s)he=OM see-ps-3sg hello do-ps-3sg
 (Whoever saw him/her said hello.)

Explaining the differences between ordinary and free RCs in Persian, Taghvaipour (2005) suggests that the complementizer *ke* is required in ordinary RCs but optional in free RCs. Sentences (3) and (4) below are instances of free RCs with and without complementizer, respectively. (DUR refers to Durative Prefix (*mi-*) (Mahootian & Gebhardt, 1997) in Persian examples and ‘modified’ under some of the Persian Examples indicates that the irrelevant parts have been deleted from the original text).

- (3) *bâ* [*hærče* *ke be dæst mi-resid*] *šen=râ* *pæxš kærd-im*.
 with whatever *ke* to hand DUR-reach-ps-3sg sand=OM spread do-ps-1pl
 (We spread the sand with whatever we found.)
- (4) [*hærki* *kæfš dâšt*] *dær mædrese ne-mi-mând*.
 whoever shoe have-ps-3sg in school Neg-DUR-stay-ps-3sg
 (Whoever had shoes on didn’t stay at school.)
 (Examples from Modir-e Madreseh ‘The School Principal’, Jalal Al-e Ahmad, 1983, pages 45 and 43, modified)

1.3 Representation of the relativized element inside RC

Comrie (1989, pp. 147–151) suggests four different possibilities for representation of the relativized element inside the modifying clause, including:

- a. **Non-reduction** in which “the head noun remains in full unreduced form in the embedded sentence in the normal position and/or with the normal case marking for a noun phrase expressing that particular function in the clause”. We use the term reiteration of the head for this case in the paper.
- b. **Pronoun retention** in which “the head noun remains in the embedded sentence in pronominal form”.
- c. **Relative pronoun** in which “there is a pronoun in the relative clause indicating the head, but instead of being in the usual position, in terms of linear

word order, for a pronoun expressing that grammatical relation, it is moved to clause initial position”.

- d. **Gap** which “simply does not provide any overt indication of the role of the head noun within the relative clause”.

In this study, we want to determine the relative frequency of each of these possible representations of the relativized element in a small-scale corpus of written Persian. It is worth mentioning that Persian ordinary RCs do not possess any relative pronouns (Windfuhr, 1979, Lazard, 1992, among others), so we skip (c).

The pronoun that Comrie points to in (b) represents resumptive (anaphoric) pronouns. RPs are pronominal variables occurring in positions in which movement has happened (McKee & McDaniel, 2001). Persian does not show a stable behavior regarding the availability of RPs in different RCs but most of them can have a personal or clitic pronoun co-indexed with the head of the relative clause. Sentences (5) and (6) below show the optionality and requiredness of RPs in Persian object and oblique RCs, respectively. (CL stands for clitic in Persian examples.)

- (5) *mærd-i [ke pâdešâh -/u=râ dær šæhr næsb kon-æd] ...*
man-DEM ke king -(s)he=OM in city appoint do-pr-3sg ...
(A man who(m) the king appoints in charge of the city ...)
(Dehkhoda Persian dictionary, 1994, under the term Šæhneh [Shahneh], modified)
- (6) *mærd-i [ke æz *-/u/æš soâl porsid-i]*
man-DEM ke from -(s)he/cl question ask-ps-2sg
(The man who you asked a questions from)

Giving a thorough description of the distribution of gaps and RPs in Persian RCs, Taghvaipour (2005) provides the following table. He also points out that the free RCs behave quite similar to restrictive ordinary RCs regarding the presence of gap or RP.

Table 1. Distribution of gaps and RPs in Persian restrictive RCs (Taghvaipour , 2005, p. 48)

	Subject	Object	Possessive*	Oblique*
GAP	Yes	Yes	No	No
RP	No	Yes	Yes	Yes

* He uses the terms Genitive and Object of Preposition RCs for Possessive and Oblique RCs, respectively.

One of the main aims of this study is to investigate the frequency of gaps and RPs in different Persian RC types.

1.4 Post-verbal RCs

Winfuhr and Perry (2010, p. 504) suggest two different orders for restrictive RCs in Persian: (i) **unmarked order** in which the RC “immediately follows the head phrase” and is embedded in the matrix sentence (7a), and (ii) **marked order** in which the RCs are “post-verbal, and have focusing-contrastive function” (7b).

- (7) (a) *ân doxtær-i [ke ali=râ dust dêr-æd] ræft.*
that girl-DEM ke Ali=OM friend have-pr-3sg go-ps-3sg
(b) *ân doxtær-i ræft [ke ali=râ dust dêr-æd].*
([that] the girl whom Ali loves left.)
(Examples from Windfuhr & Perry, 2010, p. 504, modified)

They then point to “copular RCs” (p. 505) which are necessarily post-verbal and add that this is due to the fact that the copula (linking verb) is enclitic (8).

- (8) *ân doxtær-i-st* [*ke ali=râ dust dâr-æd*]
 that girl-DEM-is ke Ali=OM friend have-pr-3sg
 (she is the girl who loves Ali.) (Windfuhr & Perry, 2010, p. 505, modified)

Thus, another characteristic of Persian RCs is that the main verb can be placed between the head DP and its modifying RC. Karimi (2001, pp. 35–36) calls these clauses “post-verbal CPs of complex DPs”. Suggesting that this phenomenon is not the result of extraposition, she claims that it is “the result of the movement of [_{DP} D N] out of the complex DP ([_{DP} D N] RC) into the Spec of a functional head” followed by the raising of the main verb “to be adjoined to the head of the focus projection”. One of the aims of this study is to see whether the Persian native speakers move [_{DP} D N] out of the complex DP or not.

2. This study

The present study is a small-scale corpus study focusing on the frequency of different RC types and different forms and features they possess.

2.1 The corpus

The corpus was composed of 1,634 sentences (i.e. 45,023 words) out of which 535 sentences contained RCs. Forty editorials of four well known newspapers (10 of each) printed in Iran including: *Donyaye Eghtesad*, *Iran*, *Keyhan*, and *Shargh* were scrutinized for the use of different RC types. The editorials had a variety of themes (political, financial, social, etc.) and were written by different writers with

different political and social views. It should be mentioned that editorials are written in formal language.

2.2 Methodology

The editorials were selected quite randomly from the online archives of these newspapers (downloaded from the website: www.magiran.com) published between June and September of 2013. All the randomly chosen articles were carefully read and scrutinized by the researchers for all kinds of RCs used. Naturally, many problems arose when the examples were analyzed. The problematic and unclear structures were considered more meticulously and discussed upon by the researchers during several meetings and discussions.

All the RCs were coded based on the following features: their general type (ordinary or free), the function of relativized element (subject, object [direct or oblique], possessive, time, or place), and whether they modified an animate or inanimate entity. Moreover, for the RCs modifying the direct object of the matrix sentence, the position of object marker *râ* (inside or after complex DP) was considered. Also, the representational form of the relativized element (gap, resumptive pronoun, pronominal clitic, or reiteration) was identified. Besides, the frequency of post-verbal RCs was determined.

3. Results and discussions

The criterion to count the sentences was the presence of a full stop, semicolon, question mark, or exclamation mark at the end of each sentence. In cases where a sentence contained more than one RC, they were counted as separate RCs. Tables 2 and 3 below present the findings.

Table 2 indicates that 32.70% (30.58 + 1.09 ordinary and 1.03 free) of all the sentences in the corpus contained RCs. The table also shows the percentage of each RC type compared to all the RCs in the corpus. The results showed that the majority of the RCs (69.53%) are ordinary subject RCs and the free RCs are the least common RC types used (2.97%). In fact the frequency of RCs goes down from subject to possessive in ordinary RCs and from subject to oblique in free RCs. This is completely in line with the Noun Phrase Accessibility Hierarchy proposed by Keenan and Comrie (1977) regarding the accessibility order of RCs based on their relativized position.

Table 2. Percentage of different RC types in the corpus

	Ordinary RCs						Free RCs		
	Subject	Object	Oblique	Possessive	Time	Place	Subject	Object	Oblique
To all sentences in corpus	22.76	3.24	2.57	2.01	0.97	0.12	0.79	0.18	0.06
	<u>30.58</u>				<u>1.09</u>		<u>1.03</u>		
To all RCs in corpus	69.53	9.90	7.85	6.16	2.99	0.37	2.42	0.37	0.18
	<u>93.44</u>				<u>3.36</u>		<u>2.97</u>		

Table 3 presents the percentage of different features (i.e. gap, RP, clitic, animate entity, head reiteration, postverbalness, and position of object marker) across the RC types. Since free RCs do not contain overt head DP, features such as, head reiteration inside RC, the post-verbal RCs, and the position of object marker *râ* could not be calculated for them. The term animate entity was used to show whether the RC refers to a [+animate] or [–animate] entity. Time and place RCs are out of the calculations for this feature, because they are always [–animate].

Table 3. Distribution of the different features across different RC types (in percentage)

RC Type		Gap	RP	Clitic	Head reiteration	Animate entity	Post-verbal RC	om ‘ <i>râ</i> ’	
								Before	After
ORC	Subject	98.11	01.61	00.26	0.00	38.44	28.76	4.30	1.07
	Object	77.35	16.98	00.00	5.66	07.54	24.52	7.54	1.88
	Oblique	07.14	90.47	00.00	2.38	04.76	38.09	0.00	0.00
	Possessive	00.00	84.85	15.15	0.00	27.27	32.00	6.06	0.00
	Time	100.0	00.00	00.00	0.00	∞	06.25	0.00	0.00
	Place	100.0	00.00	00.00	0.00	∞	00.00	0.00	0.00
FRC	Subject	100.0	00.00	00.00	∞	00.00	∞	∞	∞
	Object	66.66	33.33	00.00	∞	00.00	∞	∞	∞
	Oblique	00.00	100.0	00.00	∞	00.00	∞	∞	∞
All RC types		82.80	15.32	1.12	0.77	29.59	28.37	4.24	0.96

Notes: ∞ means that this RC type could not be analyzed for this feature

ORC stands for Ordinary RCs and FRC stands for Free RCs

Generally, RCs in this corpus were mostly used to modify inanimate entities (only 29.59% animate), which indicates a higher tendency to use RCs to modify inanimate entities in Persian written form. However, a high percentage of animate entities were modified by subject (38.44) and possessive (27.27) RCs, while no

token of free RC pointing to animate entities was observed. A few tokens of object and oblique RCs (7.54% & 4.76%, respectively) modifying animate entities were also observed.

The free RCs beginning with *wh*-words prefixed by *hær*- ‘-ever’ were not observed to be common in the corpus. Sentence (9) below is an instance of such cases. In this sentence the prefix *hær*- together with the *wh*-word *kojâ* ‘where’ act as the adjunct of place of the RC, so it can be called a place free RC. It should be mentioned that, since in our corpus we rarely encountered tokens of time and place free RCs, knowing that they could not add much to our findings, we did not count them. (Indefinite particle *-i* (IND *-i*) is used to show the indefiniteness of a noun in Persian examples)

- (9) [*hærkojâ ke bâš-æd*] *xedmætgozâr-i sâdeq bâqi xâh-æd mând*.
 wherever ke be-3sg servant-IND honest remain want-3sg remain
 (Wherever he goes he will remain an honest servant (of society).)
 (Iran, 16 July 2013, modified)

In this corpus, we faced several instances of subordinate clauses beginning with the phrase *ânče* ‘that what’ possessing all the features of free RCs. We considered them as free RCs because they were subordinate clauses acting as arguments or adjuncts in the matrix sentence and they could optionally contain the complementizer *ke*. Sentences (10) and (11) below are examples of these free RCs with and without complementizer *ke*, respectively.

- (10) [*ânče ke bær dowlæt færz bud*] *xedmæt be mærdom bud*.
 that-what ke on government mandatory was serving to nation was
 (Serving the people was what was mandatory for the government.)
 (Iran, 16 July 2013, modified)
- (11) [*ânče sâ’âti piš dær mesr ettefaq oftâd*] *tæhævvöl-i æz hæmin jens æst*.
 that-what hours ago in Egypt happen-PS-3sg transformation-IND of
 same type is
 (What happened in Egypt some hours ago was a transformation of the same type.)
 (Shargh, 4 July 2013, modified)

These sentences serve as counter-evidence for Taghvaipour’s (2005) claim that Persian free RCs necessarily begin with the prefix *hær*- ‘ever’ and a *wh*-word. However, there were some cases in which *ânče* ‘that what’ was prefixed by *hær*- ‘ever’ (12). (EZ in Persian examples stands for the *Ezafe* (meaning addition) particle referring to “an unstressed vowel *-e* (*-ye* after vowels other than *-i*) that links together elements belonging to a single constituent” (Ghameshi, 1997, p. 729) and nouns to their modifiers and possessors (Taghvaipour, 2005)).

- (12) [*hærânçe ke ræng=e hæq be xod gereft*] *mândegâr xâh-æd*
 ever-that-what ke color=EZ right to itself give-ps-3sg lasting will-3sg
šod
 become
 (Whatever was proved to be right will last forever.)

(Iran, 16 July 2013, modified)

The complementizer *ke* was present in all types of ordinary RCs, a finding which is in line with what Karimi (2001) and Mahootian and Gebhardt (1997) have proposed, but in the case of free RCs as Taghvaipour (2005) suggests, it seems to be quite optional. This claim was supported by our corpus; only 35% of the free RCs contained *ke*. (10) and (11) above are examples of the cases of free RCs with and without the complementizer *ke* detected in the corpus.

As mentioned above, *râ* is an object marker following the direct object in Persian and as Karimi (2001, pp. 3–7) suggests when it marks a DP for “specificity and accusative case” it should follow “the whole projection rather than (only) the head noun”. However, she proposes that in the case of Persian complex DPs (RC and its modified DP) acting as the object of the matrix sentence, instead of following the RC and its head, the object marker *râ* most commonly appears between the head DP and its modifying RC (13). In fact, she considers the OM following RC and its head (14) as “a recent innovation” used “in mass media and by the younger generation”, which she considers as “less acceptable”. Among all the RCs in the corpus, 5.2% (4.24 + 0.96) acted as the direct object of the matrix sentence. However, our findings confirmed Karimi’s (2001) view and the frequency of occurrence of *râ* before the RC (4.24%) in these cases was a lot more than its occurrence after the RC (.96%) which is in line with OM between the head DP and RC being more acceptable in formal Persian.

- (13) *nâpoleon qânun=e bærdeðârî=râ [ke bæd æz enqelâb=e færanse*
 Napoleon law=EZ slavery=OM ke after of revelution=EZ France
mæmnû šode bud], dobâre bærqærâr kærd.
 abolish become was, again establish do-ps-3sg
 (Napoleon re-established the law of slavery, which had been abolished after the French Revolution.) (Iran, 22 July 2013, modified)
- (14) *sokut=e æfrâd-i [ke sâl-hâ bâ išan kâr=e siyâsi*
 silence=EZ people-DEM ke year-pl with them work=EZ political
kærde-ænd]=râ xâstâr-ænd.
 do-ps-3pl=OM want-pr-3pl
 (They want the silence of the people who have cooperated with him in political affairs.) (Shargh, 25 July 2013)

As for the representation of the relativized element inside the modifying clause suggested by Comrie (1989), instances of reiteration, resumption, and gap appearing in the corpus were identified and because of the absence of relative pronouns in Persian, no token of topicalized relative pronoun was observed. The results show that 82.80% of the RCs contained gap, 16.44% resumption (15.32% RP + 1.12% clitic), and 0.77% reiteration of the head (see Table 3). As can be seen, the least common representation of the relativized element inside the RC is through reiteration of the head; an example is shown below in (15).

- (15) *inhâ=râ næ-xâstæn emtiyâz=e bozorg-i æst [ke dowlæt æz in
these=OM not-to-ask privilege=EZ great-DEM is ke government of this
emtiyaz bærxordâr æst].*
privilege possess is
(Not asking for these is a great privilege which the government possesses.)
(Iran, 16 July 2013, modified)

Disregarding a few contradictions, most of the data in this corpus were in line with Taghvaipour's (2005) proposals on the effect of function of the relativized element on the presence of gap or RP inside the RC. Despite his claim about the ungrammaticality of resumption in subject RCs, some tokens were observed containing either RP (1.61%) or pronominal clitic (0.26%). It seems that this can occur when there is a kind of emphasis on the subject of the embedded clause (16) or when the head DP is referring to a group of entities while the subject of the embedded clause is pointing to a part of the group not the whole (17–18). This is in line with Windfuhr and Perry's (2010, p. 503) suggestion that these pronouns (they use the term anaphoric pronouns for RPs) are optional in subject and object RCs and their presence indicates emphasis.

- (16) *færd=e digær-i [ke u hæm morede etehâm qærâr gereft] ...*
one=EZ another-DEM ke (s)he also accuse-of-charges was ...
(Another one who was also accused of charges ...) (Shargh, 25 July 2013)
- (17) *mævâne-i vojûd dâr-æd [ke bærxî æz ânhâ mi-tævân-æd. ræf*
obstacles-DEM there have-1sg ke some of them DUR-can-pr-3pl remove
šæv-æd].
become-pr-3s
(There are obstacles some of which can be removed.)
(Iran, 26 August 2013, modified)
- (18) *šeš pišbini-i [ke hæme-æš mohæqæq šode æst] ...*
six prediction-DEM ke all-cl come-true become is ...
(Six important predictions which all came true ...)
(Shargh, 14 August 2013, modified)

Although the optionality of gap or RP in object RC was supported by the corpus, it seems that at least in written form there is a tendency to use gaps in this RC type (77.35%).

In Persian, prepositions always precede their objects and cannot stand alone. So in oblique object RCs a personal pronoun or pronominal clitic is obligatory and gap is impossible. However, as Table 3 shows, 7.14% of oblique RCs in the corpus lack RPs or clitics. This can be accounted for by Karimi's (2001) suggestion that in cases where both the head DP and the relativized element share the same preposition, the preposition in the embedded clause can be omitted (19). All the oblique RCs lacking RP had this structure. It seems that this omission will happen only if the relativized element has the form of gap inside the RC; otherwise the preposition is obligatory before the relativized element irrespective of the two prepositions being the same or different (20).

- (19) *dær kešvæ-r-hâ-i* [ke æfrâd=e moxtælef bæ-rây-e hæ-r šoql mætræh
in country-pl-DEM ke people=EZ different for each job propose
mi-šæv-ænd
DUR-become-pr-3pl
(In countries in which different people are proposed for each job ...)
(Donyaye Eghtesad, 23 July 2013, modified)

- (20) *dær kešvæ-r-hâ-i* [ke demokrâsi dær ânhâ næhâdine šode
in country-pl-DEM ke democracy in them institutionalize become
æst]...
is
(In countries in which democracy has been institutionalized ...)
(Shargh, 4 July 2013, modified)

Unlike English, Persian does not have any relative pronoun, so there is no counterpart for “whose” in Persian. Therefore Persian possessive RCs require a personal pronoun or clitic right after the possessee. That is why most tokens of pronominal clitics in the present corpus occurred in possessive RCs (21–22). However, still the percentage was very low (15.15% only) which can be due to the formality of the texts in the corpus. It seems that these inflectional morphemes are mostly considered colloquial and Persian speakers try to avoid them in their writing.

- (21) *dowlæt-mærd-ân-i* [ke negâh-ešan be âmâr-hâ=ye eqtesâdi æst] ...
statesmen-pl-DEM ke attention-their to statistics-pl=EZ economic is
(For the statesmen whose attention is to economic statistics ...)
(Donyaye Eghtesad, 7 August 2013, modified)

- (22) *ân dowlæt=e bedehkâr [ke bedehkâri-æš titr=e ævvæl bud]*
 that government=EZ debtor ke debt-its headline=EZ first was
 (That debtor government whose debt was in the headlines ...)
 (Iran, 20 July 2013, modified)

Regarding time and place RCs, no instance of resumption was observed in the corpus. In other words, all the few witnessed tokens of time and place RCs contained gap in the position of the relativized element. Taken from the corpus, (23) and (24) below are instances of place and time RCs, respectively.

- (23) *fesâd=e mâli be jâ-i resid [ke bæxs=e xosusi hæm rošd næ-kærd].*
 corruption=EZ financial to place-DEM reach-PS ke sector=EZ private also
 develop Neg-do-ps-3sg
 (Financial corruption reached a point where the private sector didn't develop, either.)
 (Shargh, 7 July 2013, modified)
- (24) *zæmân-i [ke in forsæt=e nâder bærâye mædud-i æz æfrâd time-DEM ke this opportunity=EZ rare for few-IND of people færhæm mi-šævæd] ...*
 provision DUR-become-pr-3sg
 (A time when this rare opportunity is given to few people ...)
 (Iran, 16 July 2013-modified)

Since the frequency of free RCs in this corpus was very low (i.e. only 2.97% of all the RCs), we cannot make strong claims about their behavior. No tokens of subject free RCs containing RPs were observed in the corpus. As expected, all oblique free RCs contained RPs (25). However, about one third of object free RCs contained RPs which is in contrast with Taghvaipour's proposal regarding the ungrammaticality of resumption in object free RCs. Sentence 26 below is an example of an object free RC containing RP in the corpus.

- (25) *[ânçe æz ân goftegu mi-šæw-æd] ...*
 that-what of that discussion DUR-become-pr-3sg
 (What (that) is being discussed ...) (Shargh, 14 July 2013, modified)
- (26) *[ânçe mærdom bâ ræy=e xiš ân=râ hæmrâhi kærde-ænd]*
 that-what people with vote=EZ their that=OM accompaniment
 do-ps-3pl
 (What the people have supported with their votes)
 (Shargh, 4 August 2013, modified)

As for the post-verbal RCs, 28.37% of the ordinary RCs were separated from their modified heads by the main verbs. No token of post-verbal place RCs was observed and only 6.25% of the time RCs were post-verbal.

In this corpus, we encountered structures containing more than one RC linked by conjunction words such as *væ* 'and' modifying one head DP. There were 37 tokens of such coordinate structures in the corpus. In 81.08% of these cases, gaps occurred in all conjoined modifying RCs (27). In 10.81% of the conjoined RCs, gap occurred in the first conjunct and RP in the second one (28). In 5.40%, the opposite happened, i.e. RP appeared in the first and gap in the second (29). Also, one sentence (2.70%) was observed in which the first conjunct had RP and the second one had pronominal clitic (30). These findings were in line with Taghvaipour's (2005, p. 51) suggestion that in Persian a RP can be used with a gap in a coordinate structure and "it is possible to have gaps in both conjuncts, RPs in both, or a gap in one conjunct and a RP in the other".

- (27) *in âmâr [ke – dær mæ râkez=e motæbær=e elmi sæbt šode]*
 this statistics *ke* in centers=EZ reliable=EZ scientific register become-ps
væ [- be-âsâni qâbel-e dæstresi æst] ...
 and easily accessible is
 (These statistics which have been registered in reliable scientific centers and
 are easily accessible ...) (Keyhan, 21 July, 2013)
- (28) *čehre-hâ-i [ke – dær in sâl-hâ mæsdær=e xædæmâtî šode-ænd] væ*
 figure-pl-DEM *ke* in these year-pl source=EZ services been-3sg and
[mærdom kâr-hâ=ye ânân=râ æyân mi-bin-ænd] ...
 people work-pl=EZ they=OM clearly DUR-see-3sg
 (Those who have been the sources of some services and whose services the
 people see clearly ...) (Iran, 18 July, 2013)
- (29) *dær kešvæ-r-hâ-i [ke demokrâsi dær ânhâ næhâdine šode æst]*
 in country-pl-DEM *ke* democracy in them institutionalize become is
væ [tæeentækli-f=e extela-f-ât – be sænduq=e ræy vâgozâr mi-šævæd] ...
 and disposition=EZ difference-pl to box=EZ vote assign DUR-will-be
 (The counties in which democracy has been institutionalized and the
 disposition of disputes is left to the polls ...) (Shargh, 4 July, 2013, modified)
- (30) *tæxælof-ât=e qânun-i [ke šærh=e ân dær yâddâšt-hâ âmædeh]*
 violation-pl=EZ law-DEM *ke* explanation=EZ its in note-pl come-ps
væ [jâ=ye zekr-æš injâ ni-st] ...
 and place=EZ mention-cl-3sg here Neg-is
 (Law violations whose explanations were written in notes and whose
 mentioning is not necessary here ...) (Shargh, 11 July, 2013, modified)

Conclusion

A small-scale corpus was analyzed in this study to check the frequency and the typical syntactic characteristics of different RC types in contemporary formal written Persian. The results showed that ordinary subject RCs are more recurrent than free RCs. The results also confirmed a tendency to use object marker *râ* before the modifying RC which is in line with Karimi (2001). Besides, there were some tokens in which RPs, rather than gaps, were used in ordinary (subject) and free (object) RCs which is in contrast with Taghvaipour (2005) who points to the ungrammaticality of RPs in these RC types. There were also some instances which questioned Taghvaipour's proposal on the necessity of prefix *hær-* 'ever' in all Persian free RCs. Among three possible forms of relativized elements in Persian RCs, gap was observed to be the most common and reiteration of the head DP the least common ones. Also, some tokens were observed in support of Taghvaipour's (2005) proposals on the occurrence of gap and resumption in conjuncts of coordinate structures.

References

- Al-e Ahmad, J. (1983). *Modir-e madreseh 'The school principal'*. Tehran: Ravagh Publication.
- Comrie, B. (1989). *Language universals and linguistic typology* (2nd edition). Oxford: Basil Blackwell.
- Dabir-Moghaddam, M. (1992). On the (in)dependence of syntax and pragmatics: Evidence from the postposition-*ra* in Persian. *Cooperating with written texts: The pragmatics and comprehension of written texts* (pp. 549–573). doi: 10.1515/9783110881196.549
- Darzi, A. (2008). On the vP analysis of Persian finite control constructions. *Linguistic Inquiry*, 39(1), 103–116. doi: 10.1162/ling.2008.39.1.103
- Dehkoda, A. A. (1994). *Persian to Persian dictionary*. Tehran: University of Tehran Press.
- Ghameshi, J. (1997). Non-projecting nouns and the Ezafe construction in Persian. *Natural Language & Linguistic Theory*, 15(4), 729–788. doi: 10.1023/A:1005886709040
- Karimi, S. (2001). Persian complex DPs: how mysterious are they? *Canadian Journal of Linguistics*, 46, 63–96.
- Keenan, E., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8, 63–99.
- Lazard, G. (1992). *A grammar of contemporary Persian*. California: Mazda.
- Mahootian, Sh., & Gebhardt, L. (1997). *Persian (Descriptive grammars)*. London: Routledge. doi: 10.4324/9780203192887
- McKee, C., & McDaniel, D. (2001). Resumptive pronouns in English relative clauses. *Language Acquisition*, 9(2), 113–156. doi: 10.1207/S15327817LA0902_01
- Taghvaipour, M. A. (2005). *Persian relative clauses in head-driven phrase structure grammar*. Unpublished PhD thesis, University of Essex, England.

Windfuhr, G. (1979). *Persian grammar: history and state of its study*. New York. Mouton.

doi: 10.1515/9783110800425

Windfuhr, G., & Perry, J. R. (2010). Persian and Tajik. In G. Windfuhr (Ed.), *The Iranian Languages* (pp. 416–544). London: Routledge.

Abstract

Relative clauses in Persian: A small-scale corpus study

This study is a corpus-based investigation of Persian relative clauses (RCs) used in written mode. 535 instances of RCs occurring in 1634 sentences in 40 editorials of four newspapers published in Iran were spotted and analyzed to determine the frequency of each RC type and the occurrence of certain features including complementizer *ke*, object marker *râ*, different representations of the relativised element in the modifying clause, and the status of gap and resumptive pronoun in the RC. The results indicated that subject RCs are the most frequent types. The tendency to use object marker *râ* before the modifying RC (Karimi, 2001) was confirmed. Besides, some tokens were witnessed contrasting Taghvaipour's (2005) proposals on the ungrammaticality of resumption in Persian subject ordinary and object free RCs and the necessity of prefix *hær-* 'ever' in all Persian free RCs.

Key Words: Persian RCs, resumptive pronouns, gaps, free RCs, object marker *râ*

Authors' addresses:

Elias Abdollahnejad
School of Languages, Linguistics, Literatures
and Cultures
University of Calgary
Craigie Hall C 211
2500 University Drive NW
Calgary
Canada

elias.abdollahnejad@ucalgary.ca

Hamideh Marefat
Faculty of Foreign Languages and Literatures
University of Tehran
Between 15th and 16th St.
Kargar-e Shomali Ave.
Tehran
Iran

marefat@ut.ac.ir