



**HAL**  
open science

# A corpus-based description of cleft constructions in Persian

Pegah Faghiri, Pollet Samvelian

► **To cite this version:**

Pegah Faghiri, Pollet Samvelian. A corpus-based description of cleft constructions in Persian. *Faits de langues*, Brill, In press. halshs-02278096

**HAL Id: halshs-02278096**

**<https://halshs.archives-ouvertes.fr/halshs-02278096>**

Submitted on 4 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A corpus-based description of cleft constructions in Persian\*

Pegah Faghiri & Pollet Samvelian  
University of Cologne      Université Sorbonne nouvelle

## ABSTRACT

This paper presents a corpus-based description of cleft constructions in Persian showing that they display more diversity and complexity than currently described in the literature. Previous studies have only focused on constructions that echo one of the three main classes of clefts (IT-clefts, pseudoclefts and reversed pseudoclefts), and generally use Persian data in parallel to their English counterparts in order to contribute to the ongoing theoretical debates on the analysis of clefts. In order to achieve a more accurate picture of Persian clefts, we annotated and studied cleft and cleft-like sentences in a sample of about 550 relative clauses extracted from a journalistic corpus. Our study revealed new categories of cleft constructions that have not been reported previously; in particular, the lexically headed pseudoclefts whose usage is straightforwardly linked to the abundance of noun-verb light verb constructions in Persian. Moreover, we take issue with some claims made in prior work on the nature of the demonstrative *in* in Persian IT-clefts based on empirical arguments.

## 1. INTRODUCTION

Studies on cleft constructions in Persian have been mainly concerned with theoretical issues in the analysis of clefting in general, for which Persian data are evoked in support of a given analysis in various theoretical frameworks. These studies have established three classes of focus-marking constructions in Persian parallel to those in English and some other European languages, involving the copula *budan* and a relative clause, identified as: 1) IT-clefts; 2) pseudoclefts and 3) reversed pseudoclefts.

Although these constructions are available and attested in written corpora and in spontaneous speech, they do not seem to be the typical way of focus-marking in Persian, which is a discourse-configurational language that readily makes use of scrambling to highlight discourse prominence. Indeed, Hadian (2015) presents a quantitative corpus-based study of information structure in written Persian providing frequency data on the use of different (contrastive) focus-marking constructions and reports a very low frequency for cleft constructions (including IT-cleft and pseudoclefts), especially in comparison with scrambling, in particular left-dislocation.

In this paper, we present a small-scale corpus-study, on a sample of text from the *Hamshahri* daily, in order to identify different cleft-like patterns used in Persian to single out a focal constituent. Previous corpus studies, including only written samples as well, have focused on constructions that closely map one of the three main classes of clefts, overlooking several other cleft constructions available in Persian, including *there*-clefts. The most interesting finding of our study is however, a subtype of pseudoclefts introduced with a lexically headed relative clause in which the head noun is the nominal element of a complex predicate, ex.

---

\* We are grateful to Gerrit Dimmendaal and two anonymous reviewers for their helpful comments and suggestions on an earlier version of this paper. This work is partly supported by a public grant funded by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (reference: ANR-10-LABX-0083).

*tasmim gereftan* ‘explain’ (lit. decision take), illustrated in (1), which seem more natural in use than canonical (*wh-*) pseudoclefts.

- (1) a. *tasmim gereft ke be-rav-ad*  
 decision take.PST.3SG that SBJV-go.PRS-3SG  
 ‘(S)he decided to go.’ → Nonleft sentence
- b. *tasmim=i ke gereft in bud ke be-rav-ad*  
 decision=RSTR that take.PST.3SG that COP.PST.3SG that SBJV-go.PRS-3SG  
 ‘The decision that (s)he made was to go.’ → Lexically-headed pseudocleft.  
 (equivalent to: ‘what (s)he decided was to go’ → *wh-* pseudocleft)

Furthermore, we bring to attention several empirical issues with existing accounts, namely, with respect to the nature of the demonstrative *in*, which optionally opens Persian IT-clefts.

We will begin with some essentials about Persian information structure as well as the relative clause (section 2). We will then present an overview of previously identified cleft constructions in Persian (section 3). In section 4, we will present the findings of our corpus investigation.

## 2. PRELIMINARIES FOR UNDERSTANDING CLEFT CONSTRUCTIONS IN PERSIAN

Cleft constructions combine two syntactic constructions, namely a relative clause and a copular construction, in order to highlight a discourse prominent constituent. In what follows, we outline the properties of these two constructions in Persian. But given that clefting serves as a focus-marking device and relates to information packaging, we start by providing an outline of information structure and focus marking in Persian.

### 2.1. Information structure and focus marking

Persian is a pro-drop (null argument) language. Agreement between the verb and its subject both in person and number is (unambiguously) realized by verbal suffixes. It is also a discourse configurational language and displays free word order (Faghiri 2016). The canonical (unmarked) order of the sentence is SOV; the discourse configuration of a basic canonical sentence is given in (2). The sentential stress or nuclear accent (underlined in examples) falls, within the comment, on the focussed constituent, to be specific, on the leftmost element of the latter, as in (3).<sup>1</sup>

- (2)        Subject        Object        Verb  
               Topic        Comment

- (3)        (man) yek ketāb xarid-am  
               (I)    a        book    buy.PST-1SG  
               ‘I bought a book.’

<sup>1</sup> Glosses follow the Leipzig Glossing Rules ([www.eva.mpg.de/lingua/resources/glossing-rules.php](http://www.eva.mpg.de/lingua/resources/glossing-rules.php)). The following non-standard abbreviations are used for clarity: RA = *rā*; EZ = *Ezafe*.

Regardless of grammatical roles, the initial position is topical, while the preverbal position, carrying the primary stress, is focal, as in (4).

- (4) ketāb=rā<sup>2</sup> man xarid-am  
 book=RA I buy.PST-1SG

‘It was me who bought the book.’ / ‘As for the book, I bought it.’

Furthermore, narrow focus can always be expressed via primary stress, regardless of word order, as illustrated by the series of sentences in (5).

- (5) a. ali maryam=o busid  
 Ali Maryam=RA kiss.PST.3SG

‘Ali kissed Maryam’

What happened? Or What did Ali do?

- b. ali [<sub>F</sub> maryam=o] busid

Who did Ali kiss? (Presupposition: Ali kissed someone)

- c. [<sub>F</sub> ali] maryam=o busid

Who kissed Maryam? (Presupposition: someone kissed Maryam)

Adapted from Kahnemuyipour (2005: 170)

Leftward scrambling is also used to highlight narrow focus, in which case the focal element carries the primary stress as well, as in (6). The same is true for the expression of contrastive focus. In other words, in all sentences in (6), narrow focus can also have a contrastive reading.

- (6) a. ali barāye maryam gol xarid  
 Ali for Maryam flower buy.PST.3SG

‘Ali bought flowers for Maryam.’

- b. ali gol barāye maryam xarid

- c. gol ali barāye maryam xarid

‘Ali bought Maryam flowers (and not chocolate)’

In sum, narrow or contrastive focus can be expressed *in situ* via heavy stress, or can be expressed by scrambling, rightward for subjects and leftward for objects, combined with heavy stress.

---

<sup>2</sup> Persian exhibits differential object marking: definite and/or specific direct objects are always marked by the enclitic =*rā*, pronounced =*ro* or =*o* (after consonants) in colloquial Persian (for an overview see Samvelian 2018: 242–256).

## 2.2. Relativization

The relative clause is head external in Persian and does not involve a relative pronoun. As a subtype of subordinate clauses, it is introduced by the unique and invariable complementizer *ke*. The Relative clause always follows the modified noun, (7a). It can immediately follow the modified noun or be extraposed to the postverbal domain of the matrix clause (7b). Note that the latter option is not only permitted but also quite frequent.

- (7) a. *yek mard [ke hame=rā mi-šenāxt] vāred šod*  
 a man that all=RA IPFV-know.PST.3SG entered become.PST.3SG  
 ‘A man, who knew everybody, entered.’
- b. *yek mard vāred šod [ke hame=rā mi-šenāxt]*  
 a man entered become.PST.3SG that all=RA IPFV-know.PST.3SG  
 ‘A man entered who knew everybody.’

In restrictive relative clauses, the head noun is marked by the enclitic =*i* (see Samvelian 2006), as in (8).

- (8) *mard=i ke hame=rā mi-šenāxt vāred šod*  
 man=RSTR that all=RA IPFV-know.PST.3SG entered become.PST.3SG  
 ‘A man that knew everybody entered.’

The relativized noun can be cross-referenced by a resumptive (clitic or full) pronoun in the relative clause, as in (9). The resumptive pronoun is obligatory when the relativized function is the genitive complement of a noun (possessive) or any oblique (prepositional) function, (10) and (11) respectively.

- (9) *yek mard vāred šod ke man (u=rā) ne-mi-šenāxt-am(=aš)*  
 a man entered become.PST.3SG that I (s)he=RA NEG-IPFV-know.PST-1SG=3SG  
 (Lit.) ‘A man entered that I didn’t know (him/her).’
- (10) *mard=i ke man esm\*(=aš/=e u)=rā ne-mi-dānest-am vāred*  
 man=RSTR that I name(=3SG/=EZ<sup>3</sup> (s)he)=RA NEG-IPFV-know.PST-1SG entered  
 šod  
 become.PST.3SG  
 (Lit.) ‘A man entered that I didn’t know his/her name.’
- (11) *mard=i ke hame \*(az u/ az=aš) mi-tarsid-and vāred*

---

<sup>3</sup> Realized as an enclitic =(y)e, the ezafe links the head noun to its modifiers and to the possessor NP (Samvelian 2007).

man=RSTR that all from (s)he/from=3SG IPFV-fear.PST-3PL entered

šod

become.PST.3SG

(Lit.) ‘A man that everybody feared him/her entered.’

Free relative clauses, that is, headless relative clauses introduced by *wh*-words, like English *what*, *who(m)*, *where*, etc., are introduced by a complex pronoun formed by a *wh*- or a *wh*-like element in combination with: (i) the universal determiner *har* ‘each’, ex. *har-ke/har-kas* ‘whoever (lit. each who/each person)’, *har-če* ‘whatever’ (lit. each what) or *har-kojā/har-jā* ‘wherever’ (lit. each where/each place/), as in (12); or (ii) the demonstrative *ān*, ex. *ān-ke/ān-kas* ‘who’ (lit. that who/that person) or *ān-če* ‘what’ (lit. that what), similar to French *ce que*<sup>4</sup>, or *ān-jā* ‘where’ (lit. that place), as in (13). In these cases, the complementizer is optional (see Lazard 2006). Hereafter, for facilitation, we treat these complex pronouns as simple words without referring to their makeup.

- (12) harkas (ke) man mi-šenāxt-am      zud raft  
whoever (that) I      IPFV-know.PST-1SG early go.PST.3SG  
‘Whoever I knew left early.’

- (13) ānče (ke) goft      dorost bud  
what (that) say.PST.3SG correct COP.PST.3SG  
‘What (s)he said was right.’

In addition, it is common to use a restrictive relative clause headed by a generic noun functioning as an indefinite pronoun, such as *kas* ‘person’, *čiz* ‘thing’, as in (14a) and (14b).

- (14) a. kas=i      ke be ali salām kard=rā      ne-mi-šenāxt-am  
person=RSTR that to Ali hello do.PST.3SG=RA NEG-IPFV-know.PST-1SG  
‘I didn’t know the person who greeted Ali.’  
b. čiz=i      ke goft      dorost bud  
thing=RSTR that say.PST.3SG correct COP.PST.3SG  
‘What (s)he said was right.’

As illustrated in (15), all the three strategies can be used to form relative clauses that function as free relatives. We will see in section 3 that Persian pseudoclefts make use of the last two strategies, which are interchangeable.

- (15) a. [harkas/ānkas/kas=i      ke avval be-res-ad]      barande ast  
whoever/the one/person=RSTR that first      SBJV-arrive.PRS-3SG winner      COP.PRS-3SG

---

<sup>4</sup> Note that in Persian the choice of the pronominal element only depends on the humanness of the relativized entity and never on its function: *ke/kas* is used with human entities and *če/či/čiz* with non-human ones.

‘Whoever/Who/The one/The person that arrives first is the winner.’

- b. man [harče/ānče/čiz=i ke lāzem dār-i]=rā mi-xar-am  
I whatever/what/thing=RSTR that need have.PRS-2SG=RA IPFV-buy.PRS-1SG

‘I buy whatever/what/the thing that you need.’

### 2.3 Copular construction

Persian has an overt copula *budan*<sup>5</sup> (16), which is often realized as an enclitic in the present tense in colloquial speech, ex. (16b).

- (16) a. maryam injā bud

Maryam here COP.PST.3SG

‘Maryam was here.’

- b. (man) injā hast-am/ (man) injā=m

(I) here COP.PRS-1SG/ (I) here=COP.PRS.1SG

‘I am here.’

## 3. PERSIAN CLEFT CONSTRUCTIONS: AN OVERVIEW

Apart from Mahootian (1997), which includes a brief presentation of the main characteristics of Persian cleft and pseudocleft constructions, the latter have been the main topic of investigation in a body of studies, including Khormai and Shahbaz (2010), Moezzi-pour (2010, 2013), Khormai and Tabatabai (2012). According to these studies, Persian displays three patterns of clefting: IT-cleft, basic and reversed pseudoclefts.

### 3.1. IT-clefts

Mahootian (1997: 118) considers clefting in Persian as a case of constituent emphasis, which “moves the focused element from its unmarked position to the start of the sentence followed by a copula (...) and a *ke* ‘that’ relative clause.”

Furthermore, IT-clefts in Persian can optionally be introduced by the demonstrative pronoun *in* ‘this’. Recall that Persian is a null argument language and lacks overt expletives (Karimi 2005: 91).

- (17) (in) farhād bud ke širin=rā dust dāšt.  
this Farhad COP.PST.3SG that Shirin=RA friend have.PST.3SG

‘It was Farhad who loved Shirin.’

Moezzi-pour (2010: 182)

Accordingly, Persian IT-cleft constructions follow the pattern given in (18). The cleft clause contains the variable or the presupposition: ‘X loved Shirin’. The copular clause specifies the value of the variable, which is the focus: ‘X = Farhad’.

---

<sup>5</sup> Note that the stem is *bud-* for past tenses as in (16a), and *hast-/bāš-* for present tenses, as in (16b).

- (18) (in) +  $\underbrace{\text{clefted constituent} + \text{copula}}_{\text{copular clause}} + \text{cleft clause}$

Presupposition: *X širin=rā dust dāšt* ‘X loved Shirin’ → cleft clause

Focus: X = Farhad → copular clause

When the clefted constituent is an NP, the copula agrees with the latter in number and person (19a). However, this is not specific to cleft constructions and in binominal ordinary identification/specificational copular sentences as well the verb agrees with the second NP, as illustrated by (19b) vs. (19c), in response to the question ‘Who is Shirin’s friend?’.

- (19) a. (in) man bud-am/ \*bud            ke širin=rā dust dāšt-am.  
           this I COP.PST-1SG/ COP.PST.3SG that Shirin=RA friend have.PST-1SG  
           ‘It was me who loved Shirin.’
- b. dust=e širin farhād bud  
           friend=EZ Shirin Farhad COP.PST.3SG  
           ‘Shirin’s friend was Farhad.’
- c. dust=e širin man bud-am/        \*bud  
           friend=EZ Shirin I        COP.PST-3SG/ COP.PST.3SG  
           ‘Shirin’s friend was me.’

Various functions can be clefted: subjects, (20), direct objects, (21),<sup>6</sup> indirect and other oblique objects, (22), and adjuncts, (23) and (24).

- (20) ali=e            ke umad  
           Ali=COP.3SG that come.PST.3SG  
           ‘It is Ali who came’

---

<sup>6</sup> Note that although (21) is grammatical, we find it rather unnatural. Indeed, clefting of indefinite direct objects seems to be very marginal in Persian: there is no occurrence of such cases in our corpus study and no attested example with bare DO clefted constituents is provided in previous studies either. As already mentioned, narrow contrastive focus can be expressed through prosody and scrambling. In the case of bare direct objects, these strategies are preferred to clefting.

- (i) a. zohre barāye simā **sā'at** xarid  
           Zohreh for        Sima watch buy.PST.3SG
- b. zohre **sā'at** barā=ye simā xarid
- c. **sā'at** zohre barā=ye simā xarid  
           ‘Zohreh bought a watch for Sima (and not something else).’



Mahootian (1997: 118)

- (21) sā'at=e ke zohre xarid  
watch=COP.3SG that Zohreh buy.PST.3SG  
'It is a watch that Zohreh bought.'

Mahootian (1997: 118)

- (22) be zohre bud ke simā sā'at=o dād  
to Zohreh COP.PST.3SG that Sima watch=RA give.PST.3SG  
'It was Zohreh that Sima gave the watch to.'

Mahootian (1997: 118)

- (23) tu=ye bāq bud ke ham-digar=rā did-im  
in=EZ garden COP.PST.3SG that each-other=RA see.PST-1PL  
'It was in the garden that we saw each other.'

Mahootian (1997: 118)

- (24) diruz bud ke mehmān-hā resid-and  
yesterday COP.PST.3SG that guest-PL arrive.PST-3PL  
'It was yesterday that the guests arrived.'

Moezzi-pour (2010: 183)

But clefting adverbs is not possible, as illustrated in (25).

- (25) a. hosein mosallaman xāne=rā foruxte ast  
Hossein certainly house=RA sell.PP COP.PRS.3SG  
'Hossein has certainly sold the house.'
- b. \*(in) mosallaman ast ke hosein xāne=rā foruxte ast  
(this) certainly COP.PRS.3SG that Hossein house=RA sell.PP COP.PRS.3SG  
Intended: 'It is certainly that Hossein has sold the house.'

Khormai and Tabatabai (2012: 52)

To these functions we should also add the possibility to cleft adjectives, (26), which, to our knowledge, has not been mentioned in previous studies.

- (26) a. xām=e ke māhi zarar dār-e (na poxte)  
raw=COP.PRS.3SG that fish harm have.PRS-3SG (not cooked)

‘It is raw that fish can be harmful (not cooked).’

- b. → māhi=e xām zarar dār-e (na māhi=e poxte)  
 fish=EZ raw harm have.PRS-3SG (not fish=EZ cooked)

Non-cleft version: ‘Raw fish can be harmful (not cooked fish).’

Furthermore, in the case of prepositional objects, only the pied-piped option, that is, the pattern in which the whole PP is clefted, as in (22) above, is mentioned in previous studies. Nevertheless, similar to ordinary relativization (see section 2.2), it is also possible to cleft the NP, as in (27), in which case the preposition remains *in situ* and a resumptive pronoun is required.

- (27) (in) zohre bud ke simā beh=eš sā'at=o dād  
 (this) Zohreh COP.PST.3SG that Sima to=3SG watch=RA gave.PST.3SG

‘It was Zohreh that Sima gave the watch to (him/her).’

Cleft clauses seem to behave like non-restrictive relative clauses with respect to cross-referencing of the head noun. However, while in ordinary non-restrictive relative clauses the resumptive pronoun can either be full or enclitic, full pronouns are hardly acceptable in a cleft clause. Compare (29) and (30).

- (29) a. maryam bud ke diruz did-am=eš  
 Maryam COP.PST.3SG that yesterday see.PST-1SG=3SG  
 ‘It was Maryam that I saw yesterday.’  
 b. ?? maryam bud ke diruz u=rā did-am  
 Maryam COP.PST.3SG that yesterday (s)he=RA see.PST-1SG  
 c. ?? maryam bud ke diruz did-am  
 Maryam COP.PST.3SG that yesterday see.PST-1SG
- (30) a. maryam ke diruz did-am=eš/ u=rā did-am ...  
 Maryam that yesterday see.PST-1SG=3SG/ (s)he=RA see.PST-1SG  
 ‘Maryam, whom I saw yesterday, ...’  
 b. ??maryam ke diruz did-am ...  
 Maryam that yesterday see.PST-1SG

Finally, it is worthy to note that there seems to be a general preference for subject clefting which can explain the fact that, in case of ambiguity, for instance with a null subject relative clause, the subject reading wins (over the object reading), as illustrated by the pair of sentences in (31).<sup>7</sup> In (31a), where the subject is realized in the cleft clause, the clefted

<sup>7</sup> As we will see later in section 4.1, in all our examples of *it*-clefts, the clefted NP constituent is the subject.

constituent is straightforwardly interpreted as the direct object of the verb, cross-referenced by the resumptive enclitic. In contrast, in (31b), the cleft clause is subjectless and thus the clefted constituent is interpreted as the subject of the verb and the enclitic pronoun is interpreted as the object of the verb.<sup>8</sup> In other words, the object interpretation is excluded for the clefted constituent, *Mani*.

- (31) a. in māni<sub>i</sub> bud ke farhād diruz bord=eš<sub>i</sub> (be) park.  
 this Mani COP.PST.3SG that Farhad yesterday take.PST.3SG=3SG (to) park  
 ‘It was Mani that Farhad took to the park yesterday.’

Moezzi-pour (2010: 206)

- b. in māni<sub>i</sub> bud ke ----<sub>i</sub> diruz bord=eš<sub>j</sub> (be) park.  
 this Mani COP.PST.3SG that yesterday take.PST.3SG=3SG (to) park  
 ‘It was Mani that took him/her to the park yesterday.’

### 3.2. Basic and reversed pseudoclefts

Pseudoclefts in Persian can be construed based on two types of relative clauses discussed in Section 2.2: (i) a free relative clause introduced by a complex *wh*-pronoun formed by the demonstrative *ān*, such as *ānče* ‘what’ or *ānke* ‘the one’, (32a); or (ii) a restrictive relative clause headed by a generic noun such as *čiz* ‘thing’ or *kas* ‘person’, (32b), hereafter, *ānče* and *čiz-i ke* types respectively. Note that Moezzi-pour (2010), following Mahootian (1997), only considers the second possibility. The first possibility is however mentioned in other studies such as Khormai and Shahbaz (2010) and Khormai and Tabatabai (2012). Reversed pseudoclefts can only involve the *čiz-i ke* type relative clause, (32c).

- (32) a. ānče ke mo'arref=e nazariye=ye me'yār mi-bāš-ad in ketab  
 What that introducer=EZ theory=EZ standard IPFV-COP.PRS-3SG this book  
 ast  
 COP.PRS.3SG  
 ‘What introduces the Standard Theory is this book.’

Khormai and Shahbaz (2010: 54)

- b. kas=i ke asb dust dār-e minā=st  
 person=RSTR that horse friend have.PRS-3SG Mina=COP.PRS.3SG  
 ‘The one who likes horses is Mina.’

Mahootian (1997: 118)

<sup>8</sup> Note that in order to have an unambiguous sentence where the clefted NP maps into the pronominal object of the cleft clause, the subject must be overt.

(i) in māni<sub>i</sub> bud ke u<sub>j</sub> diruz bord=eš<sub>i</sub> (be) park.  
 this Mani COP.PST.3SG that (s)he yesterday take.PST.3SG=3SG (to) park  
 ‘It was Mani that (s)he took to the park yesterday.’

c. in ketab čiz=i ast ke mo'arref=e nazariye=ye me'yār  
 this book thing=RSTR COP.PRS.3SG that introducer=EZ theory=EZ standard  
 mi-bāš-ad  
 IPFV-COP.PRS-3SG

‘This book is what introduces Standard Theory.’

Khormai and Shahbaz (2010: 54)

The three patterns are summarized in (33). Here again, the cleft clause contains the variable or the presupposition: ‘X introduces the Standard Theory’ and the copular clause specifies the value of the variable, which is the focus: X = ‘this book’.

- (33)
- a. *ānkas/ānče/ānjā/...* + cleft clause + clefted constituent + copula  
 copular clause
- b. *kas-i/čiz-i/jā-i/...* + cleft clause + clefted constituent + copula  
 copular clause
- c. clefted constituent + *kas-i/čiz-i/jā-i/...* + copula + cleft clause  
 copular clause

Presupposition: ‘X introduces the Standard Theory’ → cleft clause

Focus: ‘X = this book’ → copular clause

### 3.3. On the nature of *in*

Recall that the demonstrative pronoun *in* ‘this’ is optional in Persian IT-clefts. Although, this should be expected given that Persian as a null-subject with a rich verbal agreement does not require expletives, it has nevertheless given rise to some discussions and claims as to the nature of *in*, as an expletive or a demonstrative (subject pronoun). Two sets of facts have been put forward:

- 1) The distribution of *in*, that is, its (in)compatibility with different types of clefted constituents (NP, PP, Adv, etc.)
- 2) Its role (or absence of role) in triggering the copula agreement.

In her analyses of *in* as a demonstrative rather than an expletive, Karimi (2005: 92), followed by Moezzi-pour (2010: 192 - 193), claims that the latter can only appear when the clefted constituent is an NP, building on constructed examples such as in (34).

- (34) a. \*in tu xiābun bud ke man did-am=eš  
 this in street COP.PST.3SG that I see.PST-1SG=3SG

Intended: ‘It was in the street that I saw him/her.’

- b. \*in be rahju bud ke man ketāb=o dād-am  
this to Rahju COP.PST.3SG that I book=RA give.PST-1SG

Intended: ‘It was to Rahjou that I gave the book.’

Karimi (2005: 92), also cited by Moezzi-pour (2010: 191)

Although we do agree with Karimi in considering *in* as a demonstrative, we should nevertheless point out that this claim: 1) is not empirically grounded, and 2) not relevant in determining the nature of *in*.

This argument is not empirically grounded given that we find attested examples of IT-clefts with a PP clefted constituent, for instance examples in (35), that sound absolutely fine to us.

- (35) a. in bārāye bār=e dovvom bud ke behtarin-hā=ye fasl=e  
this for time=EZ second COP.PST.3SG that best-PL=EZ season=EZ  
futbāl (...) mo'arrefi mi-šod-and  
football introduction IPFV-become.PST-3PL

‘It was for the second time that the best of the football season (...) was presented.’

Corpus example from Khormai and Shahbaz (2010: 58)

- b. in be ellat=e nazdiktar budan=e ān [eyvuniye] be āsiā bud ke  
this to reason=EZ closer COP.INF=EZ that [Ionia] to Aisa COP.PST.3SG that  
az tamadon-hā=ye hend va (...) čiz-hā=i gerefte bud  
from civilization-PL=EZ India and thing-PL=INDF take.PP COP.PST.3SG

‘It was because it [Ionia] was closer to Asia that it was influenced by (had taken things from) civilizations such as India and (...).

Corpus Bijankhan

However, the fact that *in* can appear with PP clefted constituents, does not rule out the analysis of *in* as a demonstrative, because a deictic *in* can appear as the subject of ordinary copular clauses with PP predicates, as in (36).

- (36) a. in barāye xune=ye mā=st  
this for house=EZ we=COP.PRS.3SG

‘This is for our house.’

- b. in tu xiābun=e mā bud  
this in street=EZ we COP.PST.3SG

‘This is in our street.’ (This happens in our street, ex. in context of photo identification)

The second issue is the function of *in*. Moezzi-pour rejects the analysis of *in* as “direct core argument of the matrix predicator” (the subject in this case), on the ground that the copula agrees in number and person with the clefted NP, and not with *in* (Moezzi-pour 2010: 193, 2013: 82).

This argument does not hold either because, as we have seen in the previous section, agreement with the deictic NP (1<sup>st</sup> or 2<sup>nd</sup> person) is ordinary in binominal copular sentences in Persian. Without wanting to discuss whether the second NP is the syntactic subject of the sentences or whether this is a case of non-canonical copular agreement, we observe that in ordinary copular sentences introduced by an anaphoric demonstrative *in*, as in (37), the copula agrees with the 1<sup>st</sup> or 2<sup>nd</sup> persons as well.

- (37) a. (in) man=am  
this I=COP.PRS.1SG  
‘This is me.’ (ex. pointing to a person on a photo)
- b. (in) to=i  
this you=COP.PRS.2SG  
‘This is you.’
- c. (in) maryam=e  
this Maryam=COP.PRS.3SG  
‘This is Maryam.’ (this person on the photo is Maryam)

Hence, contra Moezzi-pour (2010, 2013), the agreement inconsistency observed with *in* in IT-cleft sentences cannot serve as evidence to rule out *in* as a deictic element in these sentences and a core argument of the copula.

#### 4. PRELIMINARY CORPORA INVESTIGATION

Our corpus-based study is conducted on the Bijankhan corpus, a freely available corpus of more than 2.6 million tokens, extracted from the *Hamshahri* daily newspaper, manually annotated for part-of-speech information.

We manually extracted a sample of about 550 relative clauses out of the corpus. To this end, we started at the first token and stopped at around token 36 000, after extracting more than 1000 subordinate clauses. This sample included relative clauses of all possible different types. We then annotated and studied cleft and cleft-like sentences of this sample.

Overall, we have identified 31 cleft sentences, of which 18 (more or less) map one of the cleft constructions discussed in previous studies presented in section 3. These include 7 *it*-clefts, 8 basic and 3 reversed pseudoclefts. In addition, we included a case of what we call elliptic *because* IT-cleft and 5 cases of *there*-clefts. More importantly, we have identified 7 sentences that present an interesting cleft pattern overlooked in the literature. These sentences do not include a cleft clause introduced by a *wh*- or a *wh*-like element. They nevertheless present a bi-clausal pattern that, we will argue, equals that of pseudoclefts. This pattern involves noun-

verb combinations forming complex predicates, ex. *tasmim gereftan* ‘decide’ (lit. decision take).

These sentences open with a restrictive relative clause, ex. *tasmim=i ke gereft* ‘the decision that (s)he made (lit. the decision that (s)he took)’, that is comparable to a headless cleft clause, with respect to its function. Compare: *this is the decision that (s)he made* vs. *this is what (s)he decided*. Given that the verbal lexicon in Persian is mainly formed by noun-verb complex predicates,<sup>9</sup> this is a common strategy as an alternative to headless relative clauses, in particular for objects. The latter, although available, ex. *ānče tasmim gereft* ‘what (s)he decided, appears to be marginal, especially in colloquial speech.

In the remainder of this section, we first present the ordinary cleft sentences of our sample, respectively, IT-clefts, *there*-clefts, basic and reversed pseudoclefts. We then discuss sentences involving complex predicates and argue why we consider them as a type of pseudocleft.

#### 4.1. IT-clefts

Our sample contains 7 occurrences of IT-clefts. In 3 occurrences, the clefted constituent is a (pied-piped) PP, and an NP in the rest. One occurrence (38b), among the latter, includes the optional demonstrative *in*.<sup>10</sup> All NPs occupy the subject function of the cleft clause. All PPs are adjuncts.

- (38) a. az ham-in zāviye ast ke mi-guy-am (...)  
 from same-this angle COP.PRS.3SG that IPFV-say.PRS-1SG  
 ‘It is from this very angle that I am saying ...’
- b. in digar-ān hast-and ke bāyad (...) be u hediye be-dah-and  
 this other-PL COP.PRS-3PL that must (...) to (s)he gift SBJV-give.PRS-3PL  
 ‘Others are the one who should give him/her a present. (lit. It is others that should (...) give him/her a present)’

In addition, we picked out (38c) as a case of an elliptic *because* IT-cleft. This sentence opens with the copular clause *in bud ke* ‘(it) was this that’, that can be viewed as an elliptic form of *barāye in bud ke* ‘(it) was for this that’, comparable to English *It is because of this that...* Note that in Persian *barāye in ke* ‘because’ is used in answer to *why* interrogatives, see (38d). In (38c), the clefted constituent *this* is a discourse deictic that summarizes the entire preceding discourse segment, providing the reason why the person cut the tree – an event evoked earlier and thus presupposed in the discourse.

- c. in bud ke tasmim gereft deraxt=rā qa't kon-ad

<sup>9</sup> The number of simplex verbs in Persian is limited to around 250, only half of which are currently used by the speech community. The verbal lexicon is thus mainly formed of syntactic combinations, including a verb and a non-verbal element, a noun, ex. *qadam zadan* ‘walk’ (Lit. ‘step hit’), an adjective, ex. *derāz kešidan* ‘lie down’ (Lit. ‘long pull’), a particle, ex. *bar daštan* ‘take’ (Lit. ‘PARTICLE have’), or a prepositional phrase, ex. *be kār bordan* ‘use’ (Lit. ‘to work take’). These combinations are generally referred to as Complex Predicates (CPs), Compound Verbs or Light Verb Constructions (LVCs), see Samvelian (2018: 256–269) for an overview.

<sup>10</sup> Recall that we presented an example from the Bijankhan corpus of *in* with a PP clefted constituent, (35b) above. However, this example was not part of the sample we studied in our corpus study presented in this section.

this COP.PST.3SG that decision take.PST.3SG tree=RA cut do.PRS-3SG

‘This was the reason why he decided to cut the tree.’

(Lit.) ‘(It) was (for) this (reason) that he decided to cut the tree.’

d. čerā tasmim gereft deraxt=rā qa't kon-ad? barāye inke ...

why decision take.PST.3SG tree=RA cut do.PRS-3SG Because

Why did (s)he decide to cut the tree? Because...

#### 4.2. There-clefts

We have annotated 5 occurrences of presentational *there*-clefts in our sample. This cleft type, in which the whole sentence is in focus and there is no presupposed/focus articulation, is identified by Lambrecht (2001: 507-509) as sentence-focus (all focus) clefts. It displays the pattern *there is - clefted constituent - cleft clause* in English, ex. *There is [a LINGUIST] [who wants to explain CLEFTS]*, corresponding to the canonical non-cleft sentence *A linguist wants to explain clefts*.

This construction involves the verb *budan* ‘exist’ (homonymous with but different from the copula *budan*) in Persian, as in (39). The clefted constituent, occupying the subject position of the verb ‘exist’, is always an NP and can occupy different functions in the cleft clause. In our sample, we have 2 occurrences of subjects, 2 occurrences of direct objects, and 1 oblique object. In the latter case, the clefted constituent is cross-referenced by a clitic (resumptive) pronoun in the cleft sentence. Also, in one occurrence of direct object clefts the clefted constituent is cross-referenced by a full (resumptive) pronoun.

(39) a. yek goruh gozārešgar hast-and ke mi-rav-and sorāq=e

a group reporter exist.PRS-3SG that IPFV-go.PRS-3SG trace=EZ

mo'zalāt=e tehrān

problems=EZ Tehran

‘There is a group of researchers that look into Tehran’s problems.’

b. → yek goruh gozārešgar mi-rav-and sorāq=e mo'zalāt=e tehrān

a group reporter IPFV-go.PRS-3SG trace=EZ problems=EZ Tehran

Non-cleft version: ‘A group of researchers look into Tehran’s problems.’

#### 4.3. Basic and reversed pseudoclefts

We have identified 5 occurrences of pseudoclefts that closely follow one of the patterns presented in Section 3.2. There are 2 occurrences of the *ānče* type, both introduced by the complex pronoun *ānče*, (40a), and 3 occurrences of the *čiz-i ke* type, 2 with *čiz*, (40b), and 1 with *kas*, (40c).

(40) a. ānče bār=e zendegi=rā (...) sangintar mi-sāz-ad omuman

what burden=EZ life=RA heavier IPFV-make.PRS-3SG generally

ziyāderavi dar xod=e zendegi ast



immoderation in self=EZ life COP.PRS.3SG

‘What makes the burden of living heavier, is generally immoderation in life itself.’

- b. *čiz=i ke dar jāme’e ziyād ast mo'zalāt ast*  
thing=RSTR that in society much COP.PRS.3SG problem COP.PRS.3SG

‘What is abundant in the society is problems.’

- c. *dar in bačče-hā kas=i ke xub dar-āmade nimā raisi ast*  
in this child-PL person=RSTR that well PART come.PP Nima Raisi  
COP.PRS.3SG

‘Among these fellows, the one who is successful is Nima Raisi.’

We also picked out an occurrence that closely matches an ordinary *čiz-i ke* type pseudocleft, with the only exception that the cleft clause is headed by a (fully) lexical noun *maqule* ‘category’. However, in the context of the sentence, the latter is used as a lexical/stylistic variant of the generic *čiz* ‘thing’. This is even clearer if we consider the English translation.

- (41) *maqule=i ke dar barnāmesāzi=ye mā mod šode in*  
category=RSTR that in programming=EZ we fashion become.PP this  
*ast ke ...*  
COP.PRS.3SG that

‘What has become fashionable in our programming is that...’

In addition, we included two occurrences with modified indefinite pronouns *yek čiz-e jāleb* ‘an interesting thing’ (42), and *yek čiz-e digar* ‘another thing’ (43).<sup>11</sup> These sentences do not have exactly the same semantic properties as their “plain” pseudocleft counterparts. For instance, in (42) the plain pseudocleft, that is, (42b), has an exhaustive reading, while in our corpora example, (42a), this is not the case. We have nevertheless kept these sentences because they can be declefted, that is, have a mono-clausal (noncleft) corresponding sentence, with limited lexical adjustment, as in (42c) and (43c).

- (42) a. *yek čiz=e jāleb=i ke dar in barnāme bud (...) tadvin*  
one thing=EZ interesting=RSTR that in this program COP.PST.3SG editing  
*bud*  
COP.PST.3SG

‘One interesting thing that was in this program was editing.’ (non-exhaustive reading)

---

<sup>11</sup> The fact that the NP does not carry the restrictive enclitic *=i* is not relevant here.

b. *čiz=i ke dar in barnāme bud [va jāleb bud] (...)*  
*tadvin*  
 thing=RSTR that in this program COP.PST.3SG and interesting COP.PST.3SG  
 editing  
 bud  
 COP.PST.3SG  
 ‘What was interesting in this program was editing.’ (exhaustive reading)

c. *yek=i az čiz-hā=ye jāleb dar in barnāme (...) tadvin bud*  
 one=INDF from thing-PL=EZ interesting in this program editing  
 COP.PST.3SG  
 ‘One of the interesting things in this program was editing.’

(43) a. *yek čiz=e digar ke in mas'ale=rā mozā'af kard in*  
 one thing=EZ other that this problem=RA double do.PST.3SG this  
 bud ke ...  
 COP.PST.3SG that  
 ‘Another thing that intensified this problem was (that) ...’

b. *čiz=i ke [alāve bar in] in mas'ale=rā mozā'af kard in*  
 thing=RSTR that addition to this this problem=RA double do.PST.3SG this  
 bud ke ...  
 COP.PST.3SG that  
 ‘What intensified this problem, in addition to this, was (that) ...’

c. *[alāve bar in] in ke ... (ham) in mas'ale=rā mozā'af kard*  
 addition to this this that also this problem=RA double do.PST.3SG  
 ‘(In addition to this), that... (also) intensified this problem.’

In all the occurrences, the clefted constituent is an NP (simple or complex, that is, carrying a completive) and occupies the subject function in the cleft clause.

We have found 3 occurrences of reversed pseudoclefts, all involving the indefinite *čiz*. In all examples, the cleft constituent is a discourse deictic referring to an entire segment of the preceding discourse, which is characteristic of the cleft’s summative and ending role. This is indeed typical of reversed pseudoclefts and also corresponds to the statistically dominant type in other languages such as English (Collins 1991, 2005, Oberlander and Delin 1996).

Interestingly, in one example, the cleft constituent is covert (44). Recall that the cleft constituent in reversed pseudoclefts is the subject of the copular sentence and that in Persian a pronominal subject does not need to be overt.

- (44) Ø daqīqan ham-ān čiz=i ast ke man mi-xāst-am  
 exactly same-that thing=RSTR that I IPFV-want.PST-1SG  
 ‘this is exactly what I wanted. (lit. (this) is exactly the same thing that I wanted.’

In 2 occurrences the clefted constituent occupies the direct object and in 1 the oblique object function in the cleft constituent. In the latter, the clefted constituent is cross-referenced by *is* a resumptive enclitic.

#### 4.4. Lexically headed pseudoclefts involving complex predicates

If we stick to the basic definition of clefting that is, “a specific syntactic pattern which serves to separate a discourse prominent constituent structurally from the rest of the clause” (Hartmann & Veenstra 2013), we find an interesting cleft pattern involving complex predicates. In (45), we provide two (simplified) examples that involve the following complex predicates: *tozih dādan* ‘explain’ (lit. explanation give), *irād gereftan* ‘criticize’ (lit. shortcoming take), *ettefāq ofādan* ‘happen’ (lit. incident fall), (45a), (45b) and (45c) respectively.

- (45) a. tozih=i ke bāyad be-dah-am in ast ke ...<sup>12</sup>  
 explanation=RSTR that must SBJV-give.PRS-1SG this COP.PRS.3SG that  
 ‘The explanation I need to give is that ...’ (equivalent to ‘what I need to explain is that ...’)
- b. irād=i ke be mā gereft-and in bud ke ...  
 shortcoming=RSTR that to we take.PST-3PL this COP.PST.3SG that  
 ‘The shortcoming they picked out/take from us was that...’ (equivalent to ‘what they criticized us for was that ...’)
- c. ettefāq=i ke dar seri=ye qabl ofād in bud ke ...  
 incident=RSTR that in series=EZ before fall.PST.3SG this COP.PST.3SG that  
 ‘The incident that happened/fell in the previous series was that ...’  
 (equivalent to ‘what happened in the previous series was that ...’)

To illustrate the parallel between this pattern and the pseudocleft construction, let us consider (46a), an attested example of a Persian pseudocleft from Khormai and Tabatabai (2012: 45), opened by a headless (*ānče* type) cleft clause.<sup>13</sup> This example involves the complex predicate *natije gereftan* ‘deduce, conclude’ (lit. result/conclusion take).

The headless relative clause is paraphrased in (46b) by a relative clause headed by the nominal element of the complex predicate *natije*. The relative clause is underlined in both

<sup>12</sup> Note that in Persian completive predicates must be introduced by *in* and are extraposed after the copular.

(i) soāl \*(in) ast ke ...  
 question this COP.PRS.3SG that  
 ‘the question is that ...’

<sup>13</sup> We have inserted the optional complementizer *ke* in parentheses for clarification.

sentences for clarification. Note that to facilitate the understanding we use the English simple verb ‘conclude’ for (a) and the light verb construction ‘arrive at a conclusion’ for (b).

- (46) a. ānče (ke) mi-xāh-im      natije    be-gir-im      in ast  
 what (that) IPFV-want.PRS-1PL conclusion SBJV-take.PRS-1PL this COP.PRS.3SG  
 ke...  
 that  
 ‘What we want to conclude is that...’

Attested example cited in Khormai and Tabatabai (2012: 45)

- b. natij=i      ke    mi-xāh-im      be-gir-im      in ast  
 conclusion=RSTR that IPFV-want.PRS-1PL SBJV-take.PRS-1PL this COP.PRS.3SG  
 ke...  
 that  
 ‘The conclusion we want to arrive at is that...’

Let us consider English equivalents of simplified versions of these sentences (discarding the completive clause), respectively ‘what we want to conclude is this’ and ‘the conclusion we want to arrive at is this’, and compare their bi-clausal patterns:

- (a) Presupposition/variable: We want to conclude X → cleft clause  
 Focus/value: X = this → clefted constituent
- (b) Presupposition/variable: We want to arrive at the conclusion X → cleft clause  
 Focus/value: X = this → clefted constituent

We can straightforwardly see that (46b) presents the same bi-causal presupposition-focus/variable-value articulation as (46a) and can, in the same manner, map the pseudocleft construction: cleft clause (variable) – clefted constituent (value) – copula.

In order to highlight the fact that in these clefts the nominal element of the complex predicate is not used as a referential object and serves as a placeholder similar to a *wh*- element, let us consider a similar sentence involving the simplex verb *goftan* ‘say’. As illustrated by the pair of examples in (47), only in the latter case, it is possible to construct a headless pseudocleft replacing the object with a *wh*- element.

- (47) a. ānče/čiz=i      ke goft      in (jomle) bud  
 what/thing=RSTR that say.PST.3SG this sentence COP.PST.3SG  
 ‘What (s)he said was this (sentence).’
- b. \*ānče/čiz-i      ke gereft      in natije      bud  
 what/thing=RSTR that take.PST.3SG this conclusion COP.PST.3SG

Indented: ‘What (s)he concluded was this.’ (Lit. What (s)he took was this conclusion.)’

Furthermore, with the declifted and IT-cleft sentences, (48), and (49) respectively, the paraphrase where the nominal element of the complex predicate carries the variable, ex. (48a) and (49a), sounds less natural to us than the paraphrase where the variable is pronominal, ex. (48b) and (49b).<sup>14</sup>

- (48) a. *mi-xāh-im in=rā natije be-gir-im*  
 IPFV-want.PRS-1PL this=RA conclusion SBJV-take.PRS-1PL  
 ‘We want to conclude this.’
- b. *mi-xāh-im in natije=rā be-gir-im*  
 IPFV-want.PRS-1PL this conclusion=RA SBJV-take.PRS-1PL  
 ‘We want to arrive at this conclusion.’
- (49) a. *in ast ke mi-xāh-im natije be-gir-im*  
 this COP.PRS.3SG that IPFV-want.PRS-1PL conclusion SBJV-take.PRS-1PL  
 ‘It is this that we want to conclude.’
- b. *in natije=i ast ke mi-xāh-im be-gir-im*  
 this conclusion=RSTR COP.PRS.3SG that IPFV-want.PRS-1PL SBJV-take.PRS-1PL  
 ‘It is this conclusion that we want to arrive at.’

Now, we can easily show that the same is true for our corpus examples given in (45) above. In (45'), we provide the headless pseudocleft version of these sentences together with their respective variable-value articulation. Here again the variable in the noncleft corresponding sentences, given in (45''), is also best expressed by the nominal element. The completive clause is discarded in all examples for simplification.

- (45') a. *ānče (ke) bāyad tozih be-dah-am* in ast  
 what (that) must explanation SBJV-give.PRS-1SG this COP.PRS.3SG  
 ‘What I need to explain is this.’  
 Variable: I need to explain X  
 Value: X = this (+ completive)
- b. *ānče (ke) be mā (...) irād gereft-and* in bud  
 what (that) to we shortcoming take.PST-3PL this COP.PST.3SG that  
 ‘What they blamed us for (...) was this.’  
 Variable: ‘they blamed us for X

<sup>14</sup> Note that for the declifted sentences, the version without the demonstrative *in* is a possible and a better paraphrase in both cases: *mi-xāh-im natije be-gir-im ke...* ‘We want to conclude that...’.

Value: X = this (+ completive)

- c. ānče (ke) dar seri=ye qabl ettefāq oftād in bud  
what (that) in series=EZ before incident fall.PST.3SG this COP.PST.3SG

‘What happened in the previous series was this.’

Variable: x happened in the previous series

Value: X=this (+ completive)

- (45") a. bāyad in tozih=rā be-dah-am/ ? bāyad in=rā tozih be-dah-am  
must this explanation=RA SBJV-give.PRS-1SG

‘I have to explain this.’

- b. (...) in irād=rā be mā gereft-and/ ?? in=rā be mā irād gereft-and  
this error=RA to we take.PST-3PL

‘They blamed us for this.’

- c. [in ettefāq] oftād/ ?? [in] ettefāq oftād  
this incident fall.PST.3SG

‘This happened.’

To understand this construction, it is important to mention that Persian, having a limited number of simplex verbs, generally highlights a verbal argument using a nominally headed (object)<sup>15</sup> relative instead of a headless (*wh-*) relative, as in (50) and (51).<sup>16</sup>

- (50) harf=i ke piremard mi-zad...  
speech=RSTR that old-man IPFV-hit.PST.3SG

‘What the old man was saying...’

- (51) pišbini-hā=i ke ādam mi-kon-ad ma'lum nist  
prediction-PL=RSTR that one IPFV-do.PRS-3SG obvious NEG.COP.PRS.3SG

‘The predictions one does are not obvious.’ (equivalent to ‘what one predicts is not obvious.’)

---

<sup>15</sup> As we have seen in corpus examples, this possibility does not only concern object relative clauses. The nominal element can also occupy the subject position in the relative clause. In other words, there exists noun-verb complex predicates in which the noun is the subject of the verb, such as *ettefāq oftādan* ‘happen’. However, subject-verb complex predicates are limited in number in comparison to object-verb complex predicates.

<sup>16</sup> Note that a possessive construction can also paraphrase the relative clause:

- (i) harf=e piremard ...  
speech=EZ old-man  
(ii) pišbini-hā=ye ādam ma'lum nist...  
prediction-PL=EZ one obvious NEG.COP.PRS.3SG

Moreover, this pattern is quite frequently used with the simplex verb ‘have’ (*dāštan*) to express an otherwise simple possessive construction (my education: the education I have, my critique: the critique I have), as illustrated by corpus examples in (52).

- (52) a. naqd=i ke dār-am dar-mored=e ān tasvir ast  
critique=RSTR that have.PRS-1SG about=EZ that image COP.PRS.3SG  
‘The critique I have is about that image.’ (equivalent to ‘my critique is about that image.’)
- b. tahsilāt=i ke dār-am honari ast  
education=RSTR that have.PRS-1SG artistic COP.PRS.3SG  
‘The education I have is in Art.’ (equivalent to ‘my education is in Art.’)

Interestingly, the latter also gives rise to the same type of pseudocleft construction, of which we have 2 occurrences in our sample. These examples are given in (53) and (54), together with their variable/value articulation and followed by their declifted version.

- (53) a. tars=i ke dār-am in ast ke...  
fear=RSTR that have.PRS-1SG this COP.PRS.3SG that  
‘The fear I have is that ...’ (equivalent to ‘what I fear is that...’)  
Variable: I have a fear X  
Value: X=this (+ completive)
- b. → tars=e man in ast ke...  
fear=EZ I this COP.PRS.3SG that  
‘My fear is that...’
- (54) a. xāheš=i ke dār-am in ast ke...  
demand=RSTR that have.PRS-1SG this COP.PRS.3SG that  
‘The request I have is that ...’ (equivalent to ‘what I want to ask (as a favour) is that...’)  
Variable: I have a wish/request X  
Value: X=this (+ completive)
- b. → xāheš=e man in ast ke...  
demand=EZ I this COP.PRS.3SG that  
‘My request is that...’

In total, we picked out 7 lexically headed pseudoclefts. In 6 cases the head noun occupies the object position in the cleft clause, and in one the subject position. Needless to say, the clefted constituent in all cases is a complex (this + completive) NP and occupies the subject position of the copular clause.

At this point what we need to add is that this pseudocleft construction is not exclusive to Persian. If we take an English light verb construction with an equivalent simplex verb, ex. *make a decision* vs. *decide*, we can clearly see the parallelism between headless pseudoclefts and pseudoclefts headed by the nominal element of a LVC in English as well: *what I decided was to continue* vs. *the decision I made was to continue*.<sup>17</sup> The key to capture this parallelism is the fact that in these noun-verb combinations the meaning of the predicate is distributed between the noun and the verb. Hence, the nouns heading the cleft clause can also be viewed as an indefinite dummy element/placeholder/variable, in the same way as the generic (pro)noun heading a cleft clause. *Decision* in *the decision I made* is a generic (kind-level) noun denoting a subclass of nouns that are decisions, or to be more precise, the result of an act of deciding. The same is true for generic (pro)nouns *what* or (*thing*) that, while, out of context, can replace/denote anything nonhuman, in the context of the headless relative (*what I decided* or *the thing that I decided*), also denote the result of an act of deciding, a subclass of *things*, *decisions*.

Finally, it is important to note that the pattern discussed here has the same form as lexically headed pseudocleft constructions such as *The colour I prefer is red* or *The fruit I eat most is orange*. Indeed, some studies do not restrict the list of the head noun opening a pseudoclefts and include these examples as well (see De Cesare 2014 for a review). However, the pattern discussed here is structurally and semantically different. In particular, contrary to this controversial subtype that usually loses the head noun when declefted, as in *I prefer (? the colour) red* or *I mostly eat oranges*, the nominal element is always present in the declefted version, in the pattern discussed here.

#### 4.5. Summary of the results

Table 1 and Table 2 summarize the results of our corpus-study presented in detail in the three previous sections, with regards to the distribution of clefts and the nature of clefted constituent and its function (in the cleft clause), respectively. Our sample size is of course too small to allow us to drive any statistical or even meaningful quantitative conclusions. We can nevertheless make a number of observations.

1. Cleft constructions are overall infrequent. Recall that in a sample of text, totalling around 36 000 words with over 550 relative clauses, we could only extract a total of 31 cases of clefts altogether. Interesting to note that these are almost exclusively from interviews.
2. NP IT-clefts are more likely to be used to highlight focal subjects (rather than objects).
3. Headless pseudoclefts are more likely to include subject relatives while lexically headed CPr pseudoclefts include mostly object relatives.
4. Clefting of adjuncts does not seem to be common – we do not have any case of pseudoclefts and only a few cases of PP adjunct IT-clefts.
5. Reversed pseudoclefts are rare but they display typical properties of reversed pseudoclefts: they are summative and involve discourse deictic.

---

<sup>17</sup> The same is also true in French *ce que j'ai décidé* vs. *la décision que j'ai prise*.



Cleft type	Nature of the clefted constituent			Total
	NP	PP	Other	
IT-cleft	4	3	1	8
<i>There</i> -cleft	5	-		5
pseudocleft	7	-		5
Lexical pseudocleft	7	-		7
Reversed pseudocleft	2	-	1	3
<b>Total</b>	<b>23</b>	<b>5</b>	<b>2</b>	<b>31</b>

**Table 1: Nature of the clefted constituent by cleft type**

Cleft type	Function of the clefted constituent				Total
	Subject	Object	Oblique	Adjunct	
IT-cleft	4	-	-	4	8
<i>There</i> -cleft	2	2	1	-	5
pseudocleft	8	-	-	-	8
Lexical pseudocleft	1	6	-	-	7
Reversed pseudocleft	-	2	1	-	3
<b>Total</b>	<b>15</b>	<b>10</b>	<b>2</b>	<b>4</b>	<b>31</b>

**Table 2: Function of the clefted constituent by cleft type**

## 5. CONCLUSION

Exploring corpora reveals that Persian cleft constructions display more diversity and complexity than the data generally mentioned and described in theoretical studies on the syntax and semantics of Persian cleft constructions. This is not surprising since these studies have generally used Persian data in parallel to their English counterparts in order to contribute to the ongoing debates on the organization of the information structure in various types of cleft constructions. Interesting and relevant as they may be, these studies have overlooked some more specific and less canonical cleft constructions in Persian. The lexically headed pseudoclefts, for instance, whose usage is straightforwardly linked to the abundance of noun-verb complex predicates (or light verb constructions) in Persian, have never been investigated, to our knowledge, in previous studies. Our corpus-based study showed however that they constitute one of the main strategies for clefting (certain types of) direct objects in Persian.

Furthermore, although we have not carried on a quantitative study of clefting in Persian, our corpus-based study is in line with observations formulated in some previous studies (ex. Hadian 2015) as to the scarceness of clefts in (written) Persian. Future studies should focus more on the distribution of different types of cleft constructions, specially pseudoclefts and, among the latter, on lexically headed pseudoclefts.

## References

- Collins, Peter C. 1991. *Cleft and Pseudo-Cleft Constructions in English*. London: Routledge.
- Collins, Peter C. 2005. Reversed what-clefts in English Information structure and discourse function, *Australian Review of Applied Linguistics* 27/2. 63 – 74.
- De Cesare, Anna-Maria. 2014. Cleft constructions in a contrastive perspective: Towards an operational taxonomy. *Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic: Contrastive, Corpus-Based Studies*, ed. by Anna-Maria De Cesare, 9-48. Berlin: Mouton de Gruyter.
- Faghiri, Pegah. 2016. *La variation de l'ordre des constituants dans le domaine préverbal en persan : approche empirique*, (unpublished) PhD Dissertation, Université Sorbonne Nouvelle Paris 3.
- Hadian, Bahram. 2015. *Sāxtār-e ettelāi dar chand siāq-e zabān-e fārsi* [Information structure in some styles of (written) Persian]. *Proceedings of the 9th Iranian Conference on Linguistics, 1371-1394*. Tehran: Allameh Tabatabai University.
- Hartmann, Katharina and Tonjes Veenstra. 2013. Introduction. *Cleft structures*, ed. by Katharina Hartmann and Tonjes Veenstra, 1-32. Amsterdam: John Benjamins Publishing.
- Lambrecht, Knud. 2001. A Framework for the Analysis of Cleft Constructions. *Linguistics* 39/3. 463-516.
- Lazard, Gilbert, Richard, Yan, Heshmati Rokhsareh and Pollet Samvelian. 2006. *Grammaire du Persan Contemporain*. Bibliotheque iranienne. Tehran: Institut français de recherche en Iran.
- Karimi, Simin. 2005. *A Minimalist Approach to Scrambling: Evidence from Persian*. Berlin: Mouton de Gruyter.
- Kahnemuyipour Arsalan. 2004. *The syntax of sentential stress*. PhD thesis, University of Toronto.
- Khormai, Ali-Reza and Afsaneh Shahbaz. 2010. *Jomalāt-e gosaste: do maqale-ye motemāyez yā maquleyei vāhed?* [Cleft Sentences: Two Distinct Categories or One Single Category?] *Dastur [Grammar]* 5.51-81.
- Khormai, Ali-Reza and Tabatabai. 2012. *Tafāvōthā-ye goftemāni va nahvi-ye sākhtār-e gosaste va shebhe gosaste dar zabān-e fārsi* [discursive and syntactic differences between cleft and pseudocleft constructions in Persian], *Nashriye-e pazhouheshhā-ye zabpanshenāsi-ye tatbiqi [Journal of research in comparative linguistics]*1/1.33-63.
- Mahootian, Sharzad. 1997. *Persian*. New York: Routledge.
- Moezzi-pour, Farhad. 2010. *Cleft Constructions in Persian: Role and Reference Grammar Analysis*, Master thesis, University of Tehran.
- Moezzi-pour, Farhad, 2013. A constructional perspective on clefting in Persian: An insight into differentiating between emphatic and deictic in. *Linking Constructions into Functional Linguistics: The role of constructions in grammar*, ed. by Brian Nolan and Elke Diedrichsen, 67-102. Amsterdam: John Benjamins Publishing.
- Oberlander, Jon and Delin, Judy. 1996. The function and interpretation of reverse *wh*-clefts in spoken discourse. *Language and Speech* 39.185–227.

Samvelian, Pollet. 2000. L'enclitique *-i* introducteur de relative en persan : déterminant, allomorphe de l'ezāfé ou autre chose encore ? *Studia Iranica* 35/1.7–34.

Samvelian, Pollet. 2007. A (phrasal) affix analysis of the Persian ezafe. *Journal of Linguistics* 43/3.605–645.

Samvelian, Pollet. 2018. Specific features of Persian syntax. *The Oxford Handbook of Persian Linguistics*, ed. by Anousheh Sedighi and Pouneh Shabani-Jadid, 226–269. Oxford: Oxford University Press.