HAL open science

Specific Features of Persian Syntax: The Ezâfe Construction, Differential Object Marking and Complex Predictaes

Pollet Samvelian

To cite this version:

Pollet Samvelian. Specific Features of Persian Syntax: The Ezâfe Construction, Differential Object Marking and Complex Predictaes. Anousha Sedighi; Pouneh Shabani-Jadidi. Oxford Handbook of Persian Linguistics, , pp.226-269, In press, 9780198736745. halshs-01614243

HAL Id: halshs-01614243

https://shs.hal.science/halshs-01614243v1

Submitted on 27 Jun 2020

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.

Specific Features of Persian Syntax
The *Ezāfe* Construction, Differential Object Marking, and Complex Predicates

P. Samvelian

Université Sorbonne nouvelle & CNRS

1 Introduction

Three main aspects of Persian syntax have received a great deal of attention for more than 30 years: the *Ezāfe* construction, differential object marking with the enclitic =*rā*, and complex predicates. Why such enduring interest? Each of these phenomena involves language-specific challenging facts which need to be accurately described and accounted for. At the same time, each constitutes a topic of cross-linguistic investigation for which the Persian data can be of crucial interest.

The *Ezāfe* construction, a specific feature of the noun phrase in many Western Iranian languages, sheds a new light on the way dependency relationships, that is, complementation vs. modification, are realized within the noun phrase and the morphological correlates of these relationships with

respect to head vs. dependent marking patterns. It also contributes to the debate on the nature of linkers in a variety of languages.

Differential object marking (DOM) with the enclitic $=r\bar{a}$ displays a complex interaction between various semantic and discourse parameters such as referentiality, topicality, and high transitivity. Modeling the interaction between these parameters in order to account for the occurrence of $=r\bar{a}$ has been, and still is, an interesting challenge for formal and theoretical studies on Persian. Cross-linguistically, $r\bar{a}$-marking is of great interest for typological studies on DOM in the languages of the world, because of the way $=r\bar{a}$ has been grammaticalized to realize not only DOM but also topicalization, the range of grammatical functions that can be $r\bar{a}$-marked, and the role of discourse parameters.

Finally, complex predicate formation, which is the main device for enriching the verbal lexicon in Persian, provides another theoretical and typological domain of investigation in order to highlight differences and resemblances between syntactic and morphological processes of lexeme formation and the way different syntactic components contribute to the makeup of a complex lexical unit. Persian complex predicates constitute an interesting case study for theories of predicate decomposition which postulate the same underlying structure for simplex and complex predicates.

This article is devoted to these three phenomena and is divided into three sections. Each section provides an overview of empirical facts and the way various studies have tried to account for them. While it was impossible to do justice to all influential studies because of the impressive amount of work on

each topic, the article is nevertheless intended to be as exhaustive as possible and to maintain the balance between different theoretical approaches.

2 The *Ezāfe* construction

2.1 Overview and historical facts

Ezāfe, from Arabic *iḍāfa* 'addition, adjunction', designates an enclitic realized as *=(y)e*, which occurs within the noun phrase and links the head noun to its modifiers and to the possessor NP. The surface word order pattern is strongly head-initial within the Persian NP, as illustrated in ex. (1) and exemplified in ex. (2). A restricted class of determiners, quantifiers, classifiers, and adjectives precede the head noun, while all modifiers and arguments follow it. The possessor NP comes last after attributive nouns, and adjectival and prepositional modifiers. All elements occurring between the head noun and the possessor NP are linked to the head noun and to one another by the *Ezāfe*. The relative clause, on the other hand, is not introduced by the *Ezāfe* and is placed after other modifiers and the possessor NP, outside the *Ezāfe* domain. Argument prepositional phrases (PPs) are merely juxtaposed to the head noun and also occur outside the *Ezāfe* domain if the head noun is followed by either modifiers or the possessor NP, ex. (3). As shown by ex. (4), multiple modifiers may occur within the *Ezāfe* domain. In this case, the *Ezāfe* enclitic is reiterated on each modifier except the last. The possessor NP, on the other hand, is unique. In other words, if a noun has two arguments—which may be the case with eventive nouns—only one of them,

generally the second one or the Patient, can be introduced by the *Ezāfe*, ex. (5). The first argument (Agent) either is not realized or has a prepositional realization, ex. (5c).[1]

(1) (Det) N=(y)e A-N=(y)e AP=(y)e PP=(y)e NP(Poss) Rel/PP

(2) ān lebās=e arusi=e sefid=e bi āstin=e maryam
that dress=EZ wedding=EZ white=EZ without sleeve=EZ Maryam
ke pārsāl kharid
COMP last year buy.PST.3SG
'That white sleeveless wedding dress of Maryam that she bought last year'

(3) safar=e tulāni=e pārsāl=e sārā be landan
trip=EZ long=EZ last year=EZ Sara to London
'Sara's long trip to London last year'

(4) ān āvāz=e zibā=ye ghadimi=e ghamgin=e āsheghāne
that song=EZ beautiful=EZ old=EZ sad=EZ with love
'that beautiful old sad love song'

(5) a. * kharid=e maryam=e yek khāne
buying=EZ Maryam=EZ one house

b. * kharid=e yek khāne=ye maryam
buying=EZ one house=EZ Maryam

c. kharid=e yek khāne tavassot=e maryam
buying=EZ one house by=EZ Maryam
'Maryam's buying of a house'

[1] Glosses follow the Leipzig Glossing Rules (www.eva.mpg.de/lingua/resources/glossing-rules). The following non-standard abbreviations are used for clarity: CL=enclitic pronoun; COP=copula; EZ=Ezāfe; PART=particle; PFP=perfect participle; SINF=short infinitive, SPR=superlative.

The *Ezāfe* is not restricted to the NP and may also occur within adjective phrases (APs) and some PPs to link the head to its unique complement:

(6) a. āshegh=e maryam
in love=EZ Maryam
'in love with Maryam'

b. barā=ye maryam
for=EZ Maryam
'for Maryam'

The *Ezāfe* construction is not specific to Persian and is found in a significant number of Western Iranian languages (Windfuhr 1989), like Kurdish dialects (MacKenzie 1961), Hawrami (Mackenzie 1966), Zazaki (Paul 1998), and Kermanian dialects (Lecoq 2002). Although neither the shape nor the properties of the *Ezāfe* are identical from one language to another, all these languages display head-initial word order in the NP.[2] The correlation between the head-initial word order pattern and the availability of the *Ezāfe* can be accounted for on historical grounds. The enclitic *Ezāfe* has been generally assumed to have its origins in a demonstrative morpheme in Old Iranian. In Modern Persian, it can be traced back to the Old Persian relative *haya*, *hayā*, *taya* (Darmesteter 1883, Kent 1944, 1953, Meillet 1931).

Kent (1944) thoroughly argues in favor of a relative analysis of *haya*, *hayā*, *taya*, which a) in most cases introduces a subordinate clause headed by a finite verb, ex. (7a); b) takes its case from the relativized function in the subordinate clause rather than from its antecedent, accusative in ex.

[2]Note that the term "inverse" or "reverse" *Ezāfe* is sometimes used to refer to the ending—generally a vowel— occurring on pre-nominal adjectives in some Iranian languages with Adjective–Noun order, ex. Gilaki, Mazandarni, Balochi (Windfuhr 1979, pp. 27–28).

(7a), since the relativized function is the direct object in the subordinate clause. Upon 388 available instances of *haya*, *hayā*, *taya* in OP, Kent (1944) classifies 276 of them as relatives. In most of these occurrences, *haya*, *hayā*, *taya* indeed introduces a subordinate containing a finite verb (219 instances). However, Kent (1944) also groups with relatives instances such as the one in (7b), where *haya*, *hayā*, *taya* is followed only by a predicative noun phrase and the copula is lacking. The reason for this is the fact that *haya* is in the nominative case, as required by its function in the reduced relative clause (i.e., the subject of the copula), rather than in the accusative case of its antecedent NP. These copula-less constructions where *haya*, *hayā*, *taya* introduces a predicative noun phrase, ex. (7b), (7c), (7d), or an adjective phrase, ex. (7e), pave the way to its uses in Middle Persian and the emergence of the *Ezāfe* construction.[3]

(7) a. ima taya adam akunavam
this REL.NEUT.ACC 1S.NOM do.PST.1.S
'this (is that) which I did' (Kent 1944: 2, DB 1.72)

b. dārayavaum haya manā pitā
DariusACC REL.M.NOM my father.M.NOM
'Darius who (was) my father' (Kent 1944: 2, XPf 23)

[3]Haig (2011) argues *contra* Kent (1944) that the primary function of *haya*, *hayā*, *taya* was in fact to introduce appositive phrases within the NP. He draws the attention to a significant fact gone unnoticed in Kent (1944), namely the definiteness of the antecedent noun (i.e., the head noun) in most *haya* constructions. This entails that the supposed relative clause is not a restrictive relative clause (i.e. it does not contribute to the identification of the referent), but an appositive relative. Consequently, what appears to be a relative clause syntactically is functionally more of a loose appositive construction, which can in fact be extended to all uses of *haya*, *hayā*, *taya*. Determining whether *haya*, *hayā*, *taya* is an appositive "linker" rather than a relative pronoun is beyond the scope of this paper. Therefore, I will not take a stance on this issue here.

c. gaumāta haya maguš
Gaumāta.NOM REL.NOM Magian.NOM
'Gaumāta the Magian' (Kent 1944: 3, DB 1.44)

d. kāra hayā manā (...)
army this mine
'my army; the army which is mine' (Meillet 1931: §407, B. II, 87)

e. kāsaka haya kapautaka; kāsaka haya axšaina
stone REL blue; stone REL dark
'stone this blue; stone this dark'

Haya, *hayā*, *taya* becomes *ī* in Middle Persian (*haya* > *hyə* > *yə* > *ī*) and progressively loses its demonstrative/relative value to end up as a simple linker (cf. Jügel 2015: 290ff.). The possessor, as well as adjective modifiers, are introduced by the *Ezāfe* particle *ī* in Middle Persian:

(8) pus ī maz ī Ardawān
son EZ oldest EZ Artaban
'the oldest son of Artaban'

As noted by Bubenik (2009), with the consolidation of the sequence Possessee-Possessor and Modified-Modifier we reach the New Persian state of affairs. This explains why the *Ezāfe* construction is correlated with the head-initial order within the NP.

These changes in the function of the relative/linker go hand in hand with a prosodic change. Like *haya* in OP, *ī* is an independent word in Middle Persian.[4] However, given its constraint position, *ī* is generally the only element that intervenes between the head noun and the adjective or the genitive

[4]Estaji (2009) provides a set of arguments in favor of the analysis of *ī* as in independent word in MP.

modifier (Estaji 2009). This regular adjacency prepares the ground for the change of *ī* to the enclitic *=e* in New Persian.

The *Ezāfe* construction raises several issues in syntax and morphology and has thus been a particular focus of interest in numerous studies on Iranian languages (Ghomeshi 1997*a*, Haider & Zwanziger 1984, Haig 2011, Hincha 1961, Holmberg & Odden 2008, Kahnemuyipour 2014, Karimi & Brame 2012, Karimi 2007, Larson & Yamakido 2008, Palmer 1971, Samiian 1983, 1994, Samvelian 2006*b*, 2007, 2008, Schroeder 1999, among others). The most debated issue is the status of the enclitic *Ezāfe* itself and its functions. As noted by Haig (2011), the *Ezāfe* is not straightforwardly accountable in terms of available functional categories and conventional X-bar phrase structure. Narrowly related to this first issue is the internal structure of the Persian NP and the nature of dependency relationships within this syntactic domain.

The *Ezāfe* enclitic has received various and sometimes diametrically opposed analyses. It has been considered as a:

- case marker (Hashemipour 1989, Karimi 1990, Larson & Yamakido 2008, Samiian 1994);
- phonological linker, that is, an element inserted in Phonological Form, with no proper value or meaning (Ghaniabadi 2010, Ghomeshi 1997*a*, Samiian 1983);
- marker associated with the syntactic movement of the noun and realizing a strong feature (Kahnemuyipour 2014);
- linker indicating subject-predicate inversion (Den Dikken 2006);

- head-marking affix adjoined to the head noun and its intermediate projections and marking them as waiting for a dependent (Samvelian 2007, 2008).

2.2 *Ezāfe* as a case marker

The fact that the *Ezāfe* occurs as many times as there are dependents within the NP has favored its analysis as a case marker. Several studies within the generative framework have developed different variants of this analysis. Hashemipour (1989) considers the *Ezāfe* as a structural case marker on nouns, adjectives, and some PPs. For Karimi (1990), the *Ezāfe* structurally relates a head to phrases governed by the latter, by transferring the case of the head noun to its complements.

Samiian (1994) provides one of the most detailed analyses of the *Ezāfe* as a case marker. She considers the *Ezāfe* as a dummy case assigner, comparable to *of* in English, which occurs within phrases with non-case-assigning heads, that is, NPs, APs, and some PPs, and thus enables the head to case-mark its complements. Note that in this view, the *Ezāfe* structurally belongs to the modifier it precedes, while it is prosodically attached to the item it follows. The fact that the *Ezāfe* occurs in both NPs and APs is expected, since nouns and adjectives share the feature [+N], and thus do not assign case.[5] Its presence in PPs, on the other hand, is not expected, since prepositions are assumed to be case assigning. In order to overcome this problem,

[5]Samiian (1994) adopts the syntactic feature system suggested by Chomsky (1970) and subsequently developed by Jackendoff (1977), which classifies projecting lexical categories according to primitive syntactic features such as [+/- V] or [+/- N]. Adjectives and nouns are both [+N] while verbs and prepositions are [-N].

Samiian (1994) considers that those prepositions that occur with the *Ezāfe*, P2s in her classification, constitute "a kind of in-between category, sharing some properties with "true" prepositions (P1s) and some with nouns." This assumption is supported by a set of empirical facts, namely, the semantic content of P2s, their subcategorization frame, the internal structure of the PP they project—specifically their ability to allow for a specifier—and their distribution. Adopting the Neutralization Hypothesis (Riemsdijk & Williams 1981), Samiian (1994) assumes that P2s are neutralized in their [-N] feature, which leaves them with the only feature specification [-V]. Therefore, P2s cannot assign case, since only [-N] categories directly assign structural case. Sharing their only feature with the category N, they behave like the latter with respect to their case-assigning properties and need the *Ezāfe* to assign case.

While nouns, adjectives, and P2s cannot assign case, the Case Theory (Chomsky 1981) requires all NPs to receive case. The *Ezāfe* thus endorses the role of a dummy case-assigner, like *of* in English, in order to compensate for the inability of these categories to assign case. In order to account for the occurrence of the *Ezāfe* before APs, Samiian (1994) extends the case-receiving categories to include the AP. To support this idea, she refers to the case-assigning of attributive adjectives in Latin and Sanskrit.

Larson & Yamakido (2008) agree with Samiian's (1994) analysis of the *Ezāfe* as a case marker, but are not satisfied with the way it deals with modifiers. Case marking (as opposed to agreement) is typically associated with argument status; however, at least some of the *Ezāfe*-marked constituents

are modifiers. So the question arises of why modifiers should need case and what their case-assigner is. To answer these questions, Larson & Yamakido (2008) extend the shell theory of the VP (Larson 1988) to the DP. The DP is projected from the thematic structure of determiners, which assign scope and restriction as thematic roles to their arguments. For instance, the NP that combines with D saturates the quantifier restriction of the D. Under this account, (most) nominal modifiers originate as arguments of D. Therefore, they are not modifiers or adjuncts, but "oblique complements" which combine with the head prior to other arguments. All modifiers are base-generated in a post-head position. Those that bear case features, APs for instance, are required to move to a site where they can check case, that is, the pronominal position. PPs and relative clauses, on the other hand, remain *in situ*, since they do not bear case features. The fact that in Persian, and some other Iranian languages, APs also remain *in situ* is explained by the availability of the *Ezāfe*, which acts as a "generalized genitive preposition", inserted to check Case on [+N] complements of D inside the DP. Following this account, the *Ezāfe* heads its own X-bar phrase, with the modifier as complement. However, for apparently purely prosodic reasons, phonologically it attaches to the preceding item. The analysis for (9) is given in (10). The determiner *in* 'this' checks its one Case feature on its restriction. The *Ezāfe* is inserted and licenses the remaining modifiers in their base positions.

(9) in ketāb=e sabz=e jāleb
this book=EZ green=EZ interesting
'this interesting green book' (Larson & Yamakido 2008: 60, (30))

(10) Larson & Yamakido (2008: 60, (32))

[$_{DP}$ *Pro* [$_{D'}$ in [$_{DP}$ ketāb [$_{D'}$ t [$_{DP}$ [$_{XP}$ é sabz] [$_{D'}$ t [$_{XP}$ é jāleb]]]]]]]

Several studies claim that the analysis of the *Ezāfe* as a case marker faces serious problems (Ghomeshi 1997*a*, Samvelian 2007, Haig 2011). In particular, PPs headed by P1s as well as adverbial phrases do occur within the *Ezāfe* domain. However, all the above-mentioned studies acknowledge that the latter do not need to be case-marked. Moreover, the conceptual problem remains of why APs and PPs would need case. Note that in many languages, adjectives follow the noun and are neither case-marked nor introduced by any specific device.

2.3 *Ezāfe* as a phonological linker

Building on Samiian's (1983) data, Ghomeshi (1997*a*) argues against the view of the *Ezāfe* as a *morpheme* heading any sort of syntactic projection. *Ezāfe* is semantically vacuous and although it iterates throughout the NP, it is not the expression of a concord between the head noun and its dependents. Ghomeshi (1997*a*) therefore suggests viewing the *Ezāfe* not as a morpheme at all, but rather as an element inserted in Phonological Form, in a certain syntactic configuration. The need for the *Ezāfe* results from the fact that, nouns being non-projecting in Persian, a "phonological linker" must be present in order to indicate phrasing within the nominal constituent. This task is carried out by the *Ezāfe*.

The hypothesis of Persian nouns being non-projecting, which is the key-

stone of Ghomeshi's analysis, is based on a set of restrictions on the *Ezāfe* construction highlighted by Samiian (1983), (1)–(3) below, to which Ghomeshi (1997*a*) adds the restriction in (4):

1. Attributive noun phrases surface only as bare nouns in the *Ezāfe* domain, ex. (11).

2. Adjectival modifiers cannot take either nominal, ex. (12a), prepositional, ex. (12b), or sentential, ex. (12c), complements when occurring within the *Ezāfe* construction.

3. Prepositions may appear with a nominal complement within the *Ezāfe* domain, ex. (13), but sentential complements are excluded, ex. (13b).

4. NPs including a possessor are obligatorily construed as definite or presupposed, and possessors are in complementary distribution with the indefinite enclitic determiner =*i*, ex. (14).

(11) kif=e charm / * in charm
bag=EZ leather / this leather
'leather bag' / (putatively) 'a bag of this leather' (Samiian 1983: 45, (65); 53, (105))

(12) a. * mard=e negarān=e bachche-hā=yash=i vāred
man=EZ worried=EZ child-PL=CL.3.SG=INDEF entered
shod
become.PST
(intended) 'A man worried about his children entered.'

b. * mardom=e khashmgin az ertejā=ye tehrān be-pā
people=EZ angry at reactionaries=EZ Tehran to-foot
khāst-and
rise.PST-3.PL

(intended) 'The people of Tehran angry at the reactionary forces rose up.'

c. * mardom-e khoshhāl ke shāh keshvar=rā tark
people-EZ happy that Shah country=RA desertion
kard=e irān jashn gereft-and
do.PST=EZ Iran feast take.PST-3.PL

(intended) 'The people of Iran happy that the Shah left the country celebrated.' (Samiian 1983: 42, (47b), (48b), (49b))

(13) a. āftāb=e ba'd az bārun ghashang=e
sun=EZ after from rain beautiful=COP.3.SG

'The sun after the rain is beautiful.'

b. * āftāb=e ba'd az in-ke bārun bi-ād
sun=EZ after from this-that rain SBJV-come.PRS
ghashang=e
beautiful=COP.3.SG

(intended) 'The sun after it has rained is beautiful.' (Samiian 1983: 57, (124), (125))

(14) * ketāb=e sorkh=i maryam
book=EZ red=INDEF Maryam

(intended) 'a red book of Maryam'

In order to account for these facts, Ghomeshi (1997*a*) assumes that:

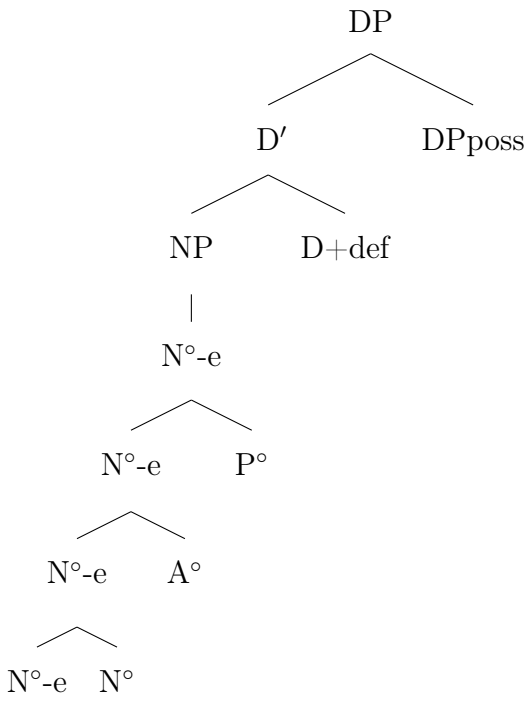
a. Persian nouns are inherently non-projecting. They never appear with filled specifier and complement positions and the NP node cannot dominate any phrasal material.

b. In spite of the fact that they are non-projecting, Persian nouns may still appear as NPs, provided they are selected by a projecting head, e.g., D°.

c. The *Ezāfe* never attaches to a phrase, which implies that the *Ezāfe* domain is the domain of X°s or bare heads.

On the basis of these assumptions, Ghomeshi (1997*a*) suggests the following structure for the Persian NP:

(15) The internal structure of the Persian NP (Ghomeshi 1997a: 780, (90))



Since Persian nouns cannot dominate phrasal material, the possessor DP, which is fully phrasal, is base-generated as sister to D′, in [Spec, DP] position. An empty D-head bearing the feature [+ def] is stipulated, whose validity is further supported by the constraint stated in (4) above.

The *Ezāfe* insertion rule, operating in PF, inserts the *Ezāfe* vowel on a lexical X° head that bears the feature [+N], when it is followed by phonetically realized, non-affixal material within the same extended projection.

In the light of this analysis, the restrictions pointed out by Samiian (1983) are straightforwardly accounted for. The only cases that would seem to resist Ghomeshi's analysis are PPs, which can occur within the *Ezāfe* domain with a complement, ex. (13a). Ghomeshi (1997*a*) claims, however, that (13a) is not a counterexample to her analysis since the noun within the modifying PP is in fact a N° and not an NP (or DP). The combination of the lexical head P° with N° provides another P° and not a P″. To support this claim, Ghomeshi (1997*a*) contrasts (16a) with (16b), imputing the ungrammaticality of the latter to the fact that the complement of the preposition *zir* 'under' is a DP containing a possessor and not an N°.

(16) a. otāgh=e kuchik=e zir=e shirvuni=e ali
room=EZ small=EZ under=EZ roof=EZ Ali
'Ali's small room under the roof'

b. * otāgh=e kuchik=e zir=e shirvuni=e jiān=e ali
room=EZ small=EZ under=EZ roof=EZ Jian-EZ Ali
(intended) 'Ali's small room, under Jian's roof' (Ghomeshi 1997a: 743, (25a-b))

This uniform analysis of modifiers as X°s in the *Ezāfe* domain has been challenged in subsequent studies by Samvelian (2007, 2008) (cf. 2.4), Ghaniabadi (2010), and Kahnemuyipour (2014) (cf. 2.5).

Ghaniabadi (2010), who adopts a similar analysis with respect to the nature of the *Ezāfe* and suggests that the *Ezāfe* is inserted by a phonological

rule in the Late-Linearization stage at PF, assumes that the modifiers occurring within the *Ezāfe* domain may either be bare heads, A°s, or phrasal, APs and PPs, and suggests the following ordering of post-nominal modifiers within the Persian NP:

(17) N A° AP PP Possessor

The main argument of Ghaniabadi (2010) for this bipartition comes from an elliptic construction he refers to as the Empty Noun Construction, where the head noun is elided leaving behind one or more modifiers. He claims that this type of ellipsis is only possible with bare adjectives, ex. (18a), and not with AP or PP modifiers, ex. (18b) and (18c) respectively. In other words, a head noun can be elided (along with other head-adjoined elements) only if the remnant is a bare adjective (A) and not an AP or a PP.

(18) a. Sajjād pirhan=e ābi pushid, Sinā ~~pirhan~~ ghermez
Sajjad shirt=EZ blue wear.PST.3SG Sina shirt red
'Sajjad wore a blue shirt, Sina a red one.' (Ghaniabadi 2010: 61)

b. * ~~keshvar-hā~~ [$_{AP}$ negarān=e afzāyesh=e gheimat=e naft]
country-PL=EZ worried=EZ increase=EZ price=EZ oil
ettelā'iyye=i sāder kard-and
statement=INDEF issue do.PST-3PL
(intended) 'The ones worried about the increase of the price of oil issued a statement.' (Ghaniabadi 2010: 69)

c. * ~~kafsh-ā~~ [$_{PP}$ tu(=ye) vitrin=e maghāze] kheili
shoe-PL inside=EZ window=EZ shop very
ghashang=e
beautiful=COP.3SG
(intended) 'The ones inside the window shop are very beautiful.'
(Ghaniabadi 2010: 70)

Samvelian (2007) and Kahnemuyipour (2014), on the other hand, argue that all post-nominal modifiers within the *Ezāfe* domain are XPs.

2.4 *Ezāfe* as a head-marking affix

Samvelian (2007, 2008) considers the *Ezāfe* as a suffix attaching to the head and to its intermediate projections in NPs, APs, and some PPs, and marking them as awaiting a modifier or a complement. Comparable to some extent to the case-marking analyses, in that the *Ezāfe* is considered a "morpheme" marking a dependency relationship between a head and its dependents, Samvelian's analysis nevertheless adopts the exact opposite standpoint in considering the *Ezāfe* as marking the head and not its dependents and forming a constituent with the head both prosodically as well as functionally. Viewed as such, the *Ezāfe* construction is an illustration of the head-marked pattern of morphological marking of grammatical relations (Nichols 1986) and reminiscent, all things being relative, of the Semitic construct state construction. This analysis entails that the *Ezāfe*, which once grouped with the constituent it introduced, has undergone a process of reanalysis-grammaticalization, being thus reinterpreted as a part of the nominal head inflection.

Samvelian (2007, 2008) builds on two sets of evidence:

a. The restrictions on the *Ezāfe* construction highlighted by Samiian (1983) and Ghomeshi (1997*a*) are either not well-grounded or are not related to the *Ezāfe per se* but to its co-occurrence with other enclitics such as the indefinite determiner =*i* or pronominal clitics.

b. The *Ezāfe*'s morphological behavior, especially its complementary distribution with the indefinite determiner =*i* and pronominal clitics, is typical of (phrasal) affixes, rather than of a post-lexical clitic (Miller 1992, Zwicky & Pullum 1983).

Example (19a) shows that an AP can be introduced by the *Ezāfe* within an NP. Note that the AP is headed by the adjective *negarān* 'worried', which shows that the ungrammaticality of (12a) above does not result from the phrasal status of the modifier headed by the adjective, but from the co-occurrence of the pronominal enclitic =*ash* and the indefinite enclitic determiner =*i*. Removing the latter makes (12a) perfectly grammatical. The same situation holds for PPs: that is, P1s, as well as P2s and P3s, can occur within the *Ezāfe* domain even when they head phrasal projections, ex. (19b) and (19c) respectively.

(19) a. abdoljalil ham bā rang=e paride va chashm-ān=e [$_{AP}$
Abdoljalil also with colour=EZ fly.PFP and eye-PL=EZ
negarān=e forurikhtan=e divār=e khāne=ash] āntaraftar
worried=EZ crumble=EZ wall=EZ house=CL.3.SG farther
istāde (...) bud
stand.PFP (...) be.PST
'Abdoljalil also with his pale figure and his eyes worried about the crumbling of the wall of his house was standing farther.' [M. Dowlatābādi, *Ruzegār-e separi shode...*]

b. ruz=e [$_{PP}$ ghabl az dastgiri=e farmānfarmā va
day=EZ before of arrest=EZ Farmānfarmā and
pesar-ān=ash]
son-PL=CL.3.SG
'the day before Farmānfarmā's and his sons' arrest' [F. Behnud,

In se zan]

c. sekke-hā=ye [$_{PP}$ tu=ye jib=e shalvār=ash]
coin-PL=EZ in=EZ pocket=EZ trousers=CL.3.SG
oftād-and
fall.PST-3.PL
'The coins in his trousers' pocket fell down.'

Samvelian (2007, 2008) then suggests accounting for the restrictions on the *Ezāfe* in morphological terms. The *Ezāfe* is viewed as a phrasal affix attaching to the head and its intermediate projections within the NP and indicating that the marked head or the intermediate projection is awaiting a dependent. Samvelian (2007) argues that the restrictions highlighted by Samiian (1983) and Ghomeshi (1997*a*) can be accounted for in terms of slot competition between the members of the class of phrasal affixes, to which belong the *Ezāfe* affix itself but also the indefinite determiner =*i* and pronominal clitics. The latter may combine with word-level inflectional affixes, that is, the plural suffix -*hā* and the definite suffix -*(h)e*, but are in complementary distribution with the members of their own class and thus mutually exclude each other. The major argument in favor of the affixal view of these enclitics is provided by restrictions on their co-occurrence: any sequence containing two or more of these enclitics is excluded, even when their scope is not the same constituent.

The incompatibility between the indefinite determiner =*i* and clitics is illustrated by (12a) above. Examples in (20), where a head noun is followed by an adjective, illustrate the same constraint on the cooccurrence between the *Ezāfe* and =*i*. Note that the indefinite determiner =*i* can either occur

to the edge of the NP, that is on the adjective, ex. (20b), in which case the modifier is introduced by the *Ezāfe*, or on the head noun, between the head noun and the adjective, ex. (20c), and in this case the *Ezāfe* is excluded, ex. (20a). These facts have led some linguists to consider that the determiner =*i* in examples such as (20c) cumulates the function of both the indefinite determiner and the *Ezāfe*. Perry (2005), for instance, uses the term "split *Ezāfe*" (p. 74) for the enclitic =*i* in these cases. Lazard (1966, p. 257) expresses a similar opinion, noting that in addition to its role as a determiner, =*i* acts in such contexts as a linker, being thus comparable to the *Ezāfe*.

(20) a. * khāne=i=e/=e=i digar
house=INDEF=EZ/EZ=INDEF another

b. khāne=ye digar=i
house=EZ another=IND

c. khāne=i digar
house=INDEF another
'another house'

Examples in (21), namely the ungrammaticality of (21b), illustrate furthermore the fact that any combination of the three enclitics under discussion is excluded even when they have different scopes. In both (21a) and (21b) a reduced relative clause (RRC), introduced by the *Ezāfe*,[6] is embedded within the NP headed by *ghahremān* 'hero'. The two NPs differ solely with respect to the constituent ordering within the reduced relative clause. In (21a), the PP *az mihan-ash* 'from his homeland' precedes the participial head of the

[6]Note that contrary to finite relative clauses, which exclude the *Ezāfe*, reduced relative clauses are always introduced by the *Ezāfe*.

modifier, while in (21b) it follows the head. Though both constituent orderings within the RRC are grammatical,[7] the addition of a possessor NP after the reduced relative is possible only in (21a) but not in (21b).

(21) a. ghahremān-e [$_{RRC}$ az mihan=ash rānde
hero=EZ from homeland=PL.3.SG drive.PFP
shode]=ye in romān
become.PFP]=EZ this novel
'the hero of this novel, (who is) driven away from his homeland'

b. * ghahremān=e [$_{RRC}$ rānde shode az
hero=EZ drive.PFP become.PFP from
mihan=ash]=e in romān
homeland=CL.3.SG=EZ this novel

Samvelian (2007) claims that this contrast can be attributed to the fact that in (21b) the *Ezāfe* is attached to the personal enclitic =*ash*, but not in (21a). Contrary to (20a), the *Ezāfe* and the personal enclitic have two different scopes in (21b): the personal enclitic is attached to the NP *mihan* 'homeland', while the scope of the *Ezāfe* is the whole N′ *ghahremān=e rānde shode az mihan=ash*.

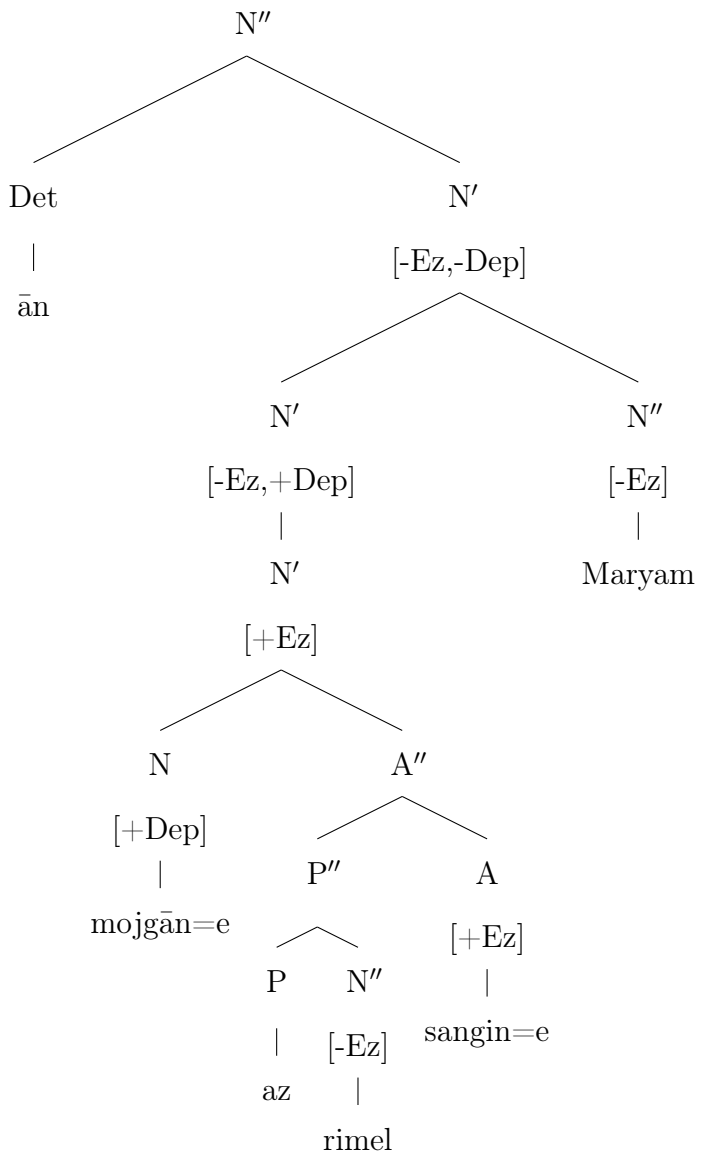
These facts are reminiscent of (haplology) phenomena discussed by Zwicky (1987) and Miller (1992), which involve the English possessive (genitive) '*s* and French weak functional words. Along the same lines of argumentation, Samvelian (2007) concludes that the *Ezāfe*, the determiner =*i*, and personal enclitics are best regarded as phrasal affixes and outlines a morphological treatment of these items in terms of edge inflection (Klavans 1985, Lapointe

[7]More generally, within APs and RRCs, the prepositional complement can either precede or follow the head, ex. *nazdik be khāne/be khāne nazdik* 'close to home'. The difference between the two orderings is a matter of register. The second ordering, where the complement precedes the head, is rather formal or literary.

1990, 1992, Tseng 2003) dealt with by word-level morphology. The *Ezāfe* is thus considered as an inflectional affix adjoining to any nominal non-maximal projection and registers the presence of a syntactic dependent, a modifier or a single NP complement, within phrases headed by a nominal category: that is, nouns, adjectives, and nominal prepositions.[8]

(22) The internal structure of the Persian NP (Samvelian 2007: 637, (57))

[8]For the details of the formalization within Head-driven Phrase Structure Grammar (HPSG) (Pollard & Sag 1994) and the way the feature [+Ez] (which indicates the presence of the *Ezāfe*) and the feature [+Dep] (which indicates that the head or the intermediate projection are awaiting a dependent) are introduced and percolated through the syntactic structure, see Samvelian (2007: 634–39).



2.5 *Ezāfe* as the result of a roll-up movement

Kahnemuyipour (2014) develops a phrasal movement analysis of the *Ezāfe* construction using what is known in the literature as roll-up movement (Cinque 2005, 2010). Contra Larson & Yamakido (2008), who take the basic word order of the Persian NP to be head-initial, Kahnemuyipour (2014)

assumes a head final ordering for Persian NPs. While in Larson and Yamakido's account, the presence of the *Ezāfe* is the result of a non-movement, in Kahnemuyipour's system, modifiers involved in the *Ezāfe* construction are uniformly merged in the specifiers of functional projections above the NP, regardless of whether they are bare or phrasal. Under this view, movement and overt morphology go hand in hand. When there is no movement, there is also no overt morphology. This implies that the pre-nominal order within the NP, i.e., the one observed in English, is the basic one. The movement that derives the postnominal order is accompanied with overt morphology, hence the existence of the *Ezāfe*, which is seen as a reflex of the roll-up movement.

The backbone of Kahnemuyipour's analysis is the near-perfect correlation between the order of the head noun and other constituents within the NP and the presence of the *Ezāfe*, with the noun clearly demarcating the distribution of the latter: *Ezāfe* cannot occur on elements surfacing before the noun and is mandatory for every element following it. Pre-nominal elements are considered as heads, i.e., X°s, and are not involved in the roll-up derivation. Therefore, they do not need the *Ezāfe*. Post-nominal elements, on the other hand, are phrases, whose surface position is the result of the roll-up derivation, leading to the appearance of the *Ezāfe* marker. A crucial aspect of this analysis is that modifiers, whether bare or phrasal, are (part of) XPs located in the specifiers of functional projections above the noun. in accordance with Bare Phrase Structure (Chomsky 1995), a bare adjective is treated as A/AP and can occupy a structural position similar to that of an AP with a complement.

Kahnemuyipour (2014) argues againt Ghaniabadi (2010), who treats bare adjectives and phrasal modifiers in radically different ways. For Ghaniabadi, bare adjectives are heads which are head-adjoined to the noun, whereas AP and PP modifiers are phrasal elements in the specifiers of functional projections above the NP. As it was mentioned in 2.3, Ghaniabadi's main argument for this bipartition comes from an elliptic construction he refers to as the "empty noun construction", where the head noun is elided leaving behind one or more modifiers. He claims that this type of ellipsis is only possible with bare adjectives, cf. ex. (18a), and not with AP or PP modifiers, cf. ex. (18b) and (18c) respectively. In other words, a head noun can be elided (along with other head-adjoined elements) only if the remnant is a bare adjective and not an AP or a PP. Ghaniabadi claims that the ellipsis of the noun along with another bare adjective is possible, because these adjectives are recursively head-adjoined to the noun. This makes it possible to elide the noun with one or more bare adjectives as long as what is left behind is another bare adjective and not an AP. Kahnemuyipour first notes that pragmatic and lexical restrictions on these elliptical constructions undermine any strong conclusion about the head vs. phrasal status of the modifiers based on the ungrammaticality of a few examples involving one or the other type of remnant. He furthermore claims that there are grammatical examples of noun ellipsis with a modifying PP and AP as remnants and provides a few examples, like the one in (23).

(23) *Context: There are two jars, one filled with wine and the other with vinegar. The speaker in contrasting the content of the jars, with serke*

'vinegar' contrastively focused and prosodically prominent.

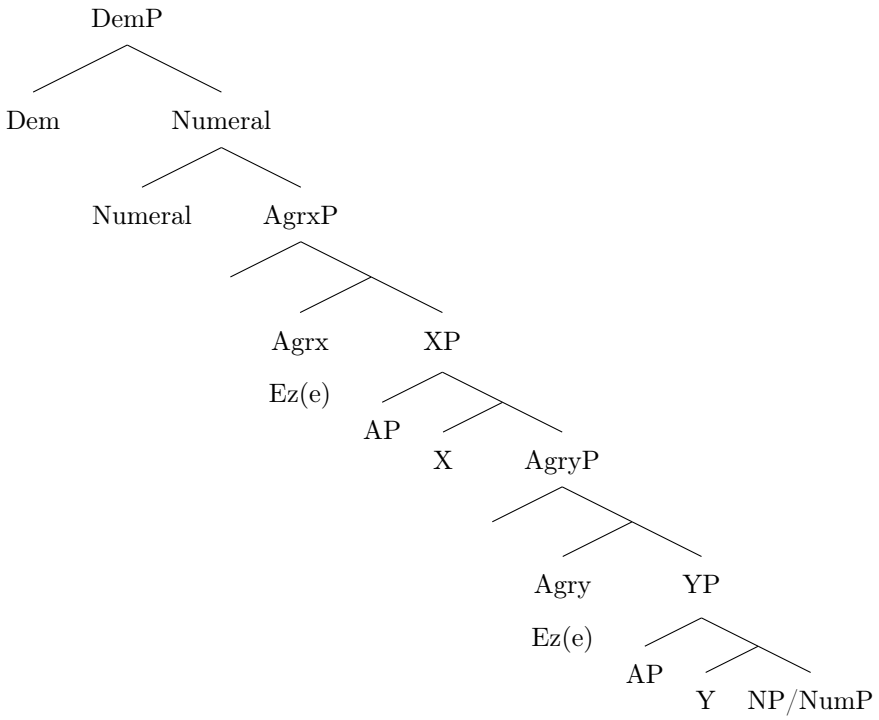
Ali tong=e [$_{AP}$ por az sharāb]=o bardāsht, ~~tong=e~~ [$_{AP}$ por az serka]=ro gozāsht
Ali jar=EZ full of wine=RA take.PST.3SG jar=EZ full of vinegar=RA leave.PST.3SG

'Ali took the jar filled with wine, and left the one filled with vinegar.'

(Kahnemuyipour 2014: 13, (25a))

Based on these facts, a uniform treatment of bare adjectives and phrasal modifiers as XPs is adopted. The Persian DP is taken to be head-final, with the NP merged at the bottom of the tree structure and the APs residing in the specifiers of projections above it. The demonstrative (Dem) and the Numeral (Num) are heads higher up in the tree structure in accordance with Cinque (2010). In addition, there are intermediate projections enabling the roll-up derivation. The relevant structures and roll-up movements are shown schematically in (24), where the projections hosting the APs are marked as XP, YP, etc. and the intermediate projections are marked as AgrPs. Under this view, the *Ezāfe* can be seen as the surface realization of the suggested inversion process, i.e., a linker in the sense of Den Dikken (2006). The height of the movement corresponds to the realization of the *Ezāfe* marker. The 'overt' movement stops below elements that are high in the universal schema such as numerals and demonstratives. Consequently, the *Ezāfe* does not occur on the latter.

(24) Deriving the *Ezāfe* construction via roll-up movement (Kahnemuyipour 2014, p. 17, (31))



2.6 Concluding remarks on the *Ezāfe* construction

As mentioned in the introductory remarks, the *Ezāfe* construction is a common feature of those Iranian languages which display a head-initial word order within their NP. While this construction is assumed to have the same origin in all these languages, it has taken a different path from one language to another, resulting in a contrasting picture in modern Iranian languages. Thus, phonological, morphological, and syntactic properties of the *Ezāfe* construction considerably cross-linguistically vary. In some languages, the *Ezāfe* particle is inflected and displays agreement features, ex. Kurmanji and Zazaki, while in others it is invariable, ex. Persian. Likewise, while it behaves rather like a post-lexical clitic in some cases, showing thus a certain degree of autonomy with respect to its host, it some other cases, it is more or less

amalgamated with other nominal inflectional affixes. The same degree of variation is observed in the syntax of the *Ezāfe* construction. While some languages treat relative finite clauses on a par with APs and PPs with respect to the *Ezāfe*, ex. Kurmanji, other languages exclude finite relatives from the *Ezāfe* construction, ex. Persian. Likewise, the order of the *Ezāfe*-marked constituents may vary: while in some languages the possessor NP closes the *Ezāfe* domain, ex. Persian, in others modifiers occurring after the possessor NP can be introduced by the *Ezāfe*, ex. Kurmanji and Sorani. Finally, the lexical head licensing the *Ezāfe* construction may also be subject to variation. While the NP is the favorite domain of the realization of the *Ezāfe* construction, adjectival and prepositional heads can also host the *Ezāfe* construction in some languages, ex. Persian, but not in all of them, ex. Sorani.

Studies on the *Ezāfe* construction have up to now focused on the description and modeling of the phenomenon in a single language, with a clear preponderance of Persian. Investigating the *Ezāfe* construction in less-studied Iranian languages using cross-linguistic approaches can shed a new light on the construction itself and on the nature of dependency relations within the NP. Another promising topic of investigation is the contact-induced *Ezāfe* construction in non-Iranian languages spoken in the era, such as Aramaic languages. Finally, broader typological studies investigating the resemblances and differences between the *Ezāfe* and linkers (loosely speaking) within the NPs in the languages of the world can also constitute another fruitful vein of research.

3 Differential object marking

Differential object marking (DOM)[9] is a rather common feature of Iranian languages and, according to Windfuhr (2009: 33), "a response to the loss of inflectional case marking". It has been one of the main topics of interest – if not the main topic – in various descriptive and formal studies of Persian, and continues to generate interest despite the significant number of publications dedicated to the topic.

In Modern Persian, DOM is realized by the enclitic *=rā*, in the formal register, and its colloquial variants *=ro* and *=o* (after a consonant). *Rā* is obligatory with all definite objects. Historically, *=rā* is the phonological reduction of *rāy* in Middle Persian, which in turn comes from the Old Persian postposition *rādi(y)*, 'for (the sake of)', 'in account of', 'concerning'. The suffix *=rā* developed as an indirect object marker in late Middle Persian and Early New Persian and progressively developed into a direct object marker only in the course of several centuries.[10] According to Paul (2003, p. 182), unlike in later classical and Modern Persian, it is not predominantly definiteness that determines when *=rā* occurs and when not, but animacy.

The enduring interest in *=rā* is due to the fact that a cluster of heterogeneous parameters seem to be at work in *rā*-marking, since *=rā* can also occur with indefinite direct objects and even other grammatical functions. Spotting *the* relevant or prevailing parameter(s) that determine(s) the presence

[9]The term was coined by Bossong (1985), who provides a detailed account of the phenomenon in a variety of languages.

[10]For a detailed discussion on *=rāy* in Middle Persian see Jügel (2015: 192–218, 340–342).

of =*rā* has thus been the major issue in studies on DOM in Persian. Another issue that has been extensively investigated in generative studies is the syntactic consequence of *rā*-marking and whether *rā*-marked objects occupy a different syntactic position with respect to their non-marked counterparts.

3.1 *Rā* as a mark of specificity

Cross-linguistic studies on DOM (Aissen 2003, Bossong 1985, Comrie 1979, 1989, Dalrymple & Nikolaeva 2011, Hopper & Thompson 1980, Lazard 1982, 1984, Malchukov 2008, Næss 2004, 2007, de Swart 2007, among others) have shown that animacy and definiteness (or specificity) are generally involved in DOM: animate and/or definite objects are more likely to be marked than inanimate and/or indefinite objects.

Among these two semantic properties acting upon DOM cross-linguistically, the degree of determination (i.e., definiteness or specificity) prevails in Persian. Grammars and linguistic studies generally qualify =*rā* as the mark of the definite direct object (Gharib et al. 1994, Khānlari 1972, Lazard 1957, Mahootian 1997, Sadeghi 1970, among many others). All definite direct objects must be *rā*-marked. This implies that personal pronouns, proper nouns, and NPs introduced by a definite determiner (e.g. demonstrative, interrogative) must be marked, ex. (25a). This also implies that all definite descriptions and NPs whose reference is unique, ex. (25b), anaphoric NPs, and all those NPs whose reference is given by the context must also be followed by =*rā* too, ex. (25c). The omission of =*rā* in all these cases yields strict ungrammaticality.

(25) a. shomā=rā/ maryam=rā/ ān dokhtar=rā did-am
you=RA/ Maryam=RA/ that girl=RA see.PST.1SG
'I saw you/ Maryam/ that girl.'

b. nakhost vazir=e farānse=rā did-am
prime minister=EZ France=RA see.PST-1SG
'I saw the French prime minister.'

c. ketāb=rā kharid-i?
book=RA buy.PST-2SG
'Did you buy the book?'

However, although =*rā* is absolutely required with definite NPs, definiteness cannot account for the whole range of the distribution of =*rā*. In other words, while definiteness constitutes a sufficient condition for *rā*-marking of DOs, it is not a necessary condition, since indefinite objects may as well be *rā*-marked, ex. (26).

(26) maryam zan=i(=rā) dar kuche did
Maryam woman=INDEF(=RA) in street see.PST-3SG
'Maryam saw a woman in the street.'

Note that in this latter case, the omission of =*rā* does not render the sentence ungrammatical.

Some authors have therefore suggested that specificity, rather than definiteness, is responsible for *rā*-marking (Browne 1970, Browning & Karimi 1994, Karimi 1990, 1996). In this view, the occurrence of =*rā* with indefinite NPs is not optional, but depends on the reading of the NP: all specific objects must be *rā*-marked, be they definite or indefinite. Note that there is no consensual definition of the notion of specificity. Informally speaking, the referent of a specific indefinite expression is identifiable to the speaker (but

not to the addressee). A prototypical specific indefinite is generally assumed to have wide scope, a referential reading, and an existential presupposition.[11] Karimi (1996) suggests that a specific NP must be *rā*-marked if it occurs in the syntactic configuration, in (27):[12]

(27) [CP ... NP... [β ... α ...] ...]

In contrast with this categoric view, other studies insist on the fact that a cluster of features or properties, and not a single binary feature (be it definiteness or specificity), is involved in DOM. Lazard (1982, 1994) claims that, apart from definiteness, the presence of =*rā* can be triggered by factors such as animacy (or humanness), the semantic "contentfullness" of the verb, the semantic "distance" between the verb and the object, the relative weight of the syntactic constituents, and finally the information structure. Lazard's approach combines thus a cluster of non-homogeneous parameters, involving not only the inherent semantic properties of the object itself but also its relationship with the verb and particularly the way speakers organize their utterance in order to "polarize" the object. Lazard coins the term "polarized object" to designate *rā*-marked objects, as opposed to "depolarized", i.e., non-*rā*-marked, objects. Because of the complex interaction between these factors, he concludes that it is impossible to formulate a categoric rule for *rā*-marking in Persian. Note that Bossong (1991) also makes the same remark on DOM in variety of languages,[13] for example, Hindi, Kannada, and Ostyak,

[11]For a detailed discussion of the notion and controversies on specificity, see Fodor & Sag (1982), Hintikka (1986), Enç (1991), and Farkas (1995, 2002).

[12]S. Karimi suggests a different view of *rā*-marking in a recent work (Karimi & Smith 2015). See also S. Karimi in this volume.

[13]For a detailed discussion on this point see also Dalrymple & Nikolaeva (2011).

and claims that the rules of DOM in these languages must allow for a certain degree of variability across speakers and situations.

The fact that specificity is not sufficient by itself to account for the whole range of the uses of =*rā* has been noted by several other authors (Dabir-Moghaddam 1992, Ghomeshi 1997*b*, Meunier & Samvelian 1997, among others). The most obvious counterexample to such a generalization is provided by the use of =*rā* with generic objects, ex. (28):

(28) a. jāni=rā mojāzāt mi-kon-and
murderer=RA punishment IPFV-do.PRS-3PL
'Murderers are punished.' (lit. 'They punish murderers.') (Lazard 1982: (43))

b. serke shir=rā mi-bor-ad
vinegar milk=RA IPFV-cut.PRS-3SG
'Vinegar curdles milk.' (adapted from Phillott 1919: 455)

c. mi-dān-id chetor gusfand=rā mi-kosh-and
IPFV-know.PRS-2PL how sheep=RĀ IPFV-read.PRS-3PL
'Do you know how sheep are killed?' (lit. Do you know how they kill sheep?' (adapted from Phillott 1919: 459)

d. arusi-hā=rā injā zohr-hā mi-gir-and
wedding-PL=RA here noon-PL IPFV-catch.PRS-3PL
'Weddings are celebrated here at noon.' (lit. 'Here, they celebrate weddings at noon.') (M. Sādeghi, *Kalle-ye asb*)

Phillott (1919) notes that the omission of =*rā* in (28c) would not change the interpretation of the sentence. A very convincing example in this sense is provided by Hincha (1961), quoted by Lazard (1982):

(29) arabi balad=i? (...) torki=rā balad=i?
Arabic knowing=COP.2SG Turkish=RA knowing=COP.2SG

'Do you know Arabic? And what about Turkish, do you know Turkish?'

As Lazard (1982) notes, (29) uncontroversially proves that referentiality is not the only trigger of DOM in Persian and in some cases does not play any role at all.

3.2 *Rā* as a mark of high transitivity

Another set of data showing that other factors than referentiality intervene in *rā*-marking is provided by the following examples:

(30) a. ru=ye yakh=rā kāh mi-pash-and
on=EZ ice=RA straw IPFV-spread.PRS-3PL
'[They] spread straw on the ice.' (Lazard 1982: (63))

b. zamin=e posht=e bāgh=rā lubiyā kāsht-am
ground=EZ behind=EZ garden=RA bean plant.PST-1SG
'I have planted beans in the ground behind the garden.' (A. M. Afghāni, *Butezār*)

yeki digar mi-goft injā=rā behtar ast yek
one other IPFV-say.PST.3SG here=RĀ better be.PRS.3SG one
tappe be-kesh-id
hill SBJV-draw.PRS.2PL
'Another one was saying: here, it would be better to draw a hill.'

(B. Takhti, *Kāsh nām-ash rā miporsidi*)

c. give=rā na-bāiad nakh bast
espadrille=RA NEG-must string attach.SINF
'One should not attach laces on espadrilles.' (M. Dowlatābādi, *Jāi-e xāli-e Soluch*)

d. shāiad u=rā ham āzār-hā resānde bud-and
Perhaps he=RA also persecution-PL convey.PFP be.PST-3PL
'Perhaps they have also harmed him.' (F. Sāri, *Morvārid Xātun*)

e. masalan be jā=ye pānsad nafar, emshab hezār
For example instead=EZ five hundred person tonight thousand
nafar=rā dar masjed=e āzarbāijāni-hā shām
person=RA in mosque=EZ Azerbaijani-PL dinner
be-dah-ad
SBJV-give.PRS.3SG
'For example, instead of 500 people, tonight, he'd better offer dinner to 1000 people at the Azerbaijanis' mosque.' (M. Ettehādiye, *Zendegi bāiad kard*)

In these examples, the *rā*-marked constituent is not the DO properly speaking, but a locative or a "dative" argument of the verb, which can also have a prepositional realization. In example (30b), for instance, *zamin=e posht=e bāgh* 'the ground behind the garden' is by preference introduced by a locative preposition such as *dar* 'in', as illustrated in example (31a). In example (30e), *hezār nafar* 'one thousand people', which is the goal or beneficiary argument of the verb *dādan*, is canonically introduced by the preposition *be* 'to', ex. (31b).

(31) a. dar zamin=e posht=e bāgh lubiyā kāsht-am
in ground=EZ behind=EZ garden bean plant.PST-1SG
'I have planted beans in the ground behind the garden.'

b. be hezār nafar shām mi-dah-ad
to one thousand person dinner IPFV-give.PRS.3SG
'He offers dinner to one thousand people.'

Lazard (1982) claims that in (30a), the whole surface or space designated by the Ground argument, i.e. *ru=ye yax* 'on the ice', is occupied by the

result of the action. In other words, the *rā*-marked variant of the locative argument implies a holistic reading, while the prepositional variant gives only a locative indication. Therefore, (30b) and (31a) do not display the same truth conditions. Interestingly, the same remarks have been made for the *spray-load* alternation in English (Levin 1993), *to spray paint on the wall* vs. *spray the wall with paint*. It has been claimed that in this latter case, the locative argument receives a holistic reading (Anderson 1971).

The alternation illustrated by (30e) and (31b) on the other hand is comparable to the so-called "dative alternation" in English (*to give something to somebody* vs. *to give somebody something*). The interaction between several parameters, such as definiteness, animacy, and discourse accessibility (or givenness), has been shown to favor the double object variant (Bresnan et al. 2007). It seems at first sight that some of these parameters also play a role in the preference for the double object construction in Persian, although the phenomenon needs to be thoroughly investigated.

Lazard (1982) coins the term "Polarized Quasi-Objects" for the *rā*-marked constituents in (30). The role of *rā* is thus to turn some oblique arguments, i.e. those that share some typical properties of *rā*-marked DOs, into objects. The use of =*rā* in this function is not limited to locative (or Ground) and "dative" (Goal) arguments but extends to a wide variety of cases, illustrated by the following examples:

(32) a. be khodā man hāzer=am sad farsakh rāh=ro bā
to God I ready=COP.1SG hundred league path=RA with
to piāde bi-ā-m
you on foot SBJV-come.PRS-1SG
'Believe me, I'm ready to walk one hundred leagues with you.'

(lit. 'I'm ready to come one hundred leagues with you.')

b. zohr, nāhār=rā bā yek dust (...) dar restorān=e farid
noon, lunch=RA with a friend (...) in restaurant=EZ Farid
(...) gharār o madār dār-am
(...) appointment have.PRS-1SG
'At noon, for lunch, I have an appointment with a friend at Farid restaurant.' (E. Fasih, *Pārs*)

c. shab=rā dar ghahvekhāne manzel kard-im
night=RA in cafe stop-off do.PST-1PL
'At night, we stayed at the café.' or 'We spent the night at the café.' (Lazard 1982: (70))

d. man=e bichāre che be-gui-am ke
I=EZ miserable what SBJV-say.PRS-1SG that
zendegi=am=rā ranj bord-am (...)
life=CL.1SG=RA pain carry.PST-1SG
'What can I say, I who am miserable and who have suffered all my life (...).' (Lazard 1982: (71))

In (32a), the *rā*-marked constituent is a modifier denoting the distance. In (32b), =*rā* occurs with an adjunct expressing the purpose. In (32c) and (32d), =*rā* is adjoined to temporal modifiers.

Concerning the presence of =*rā* with temporal modifiers, Lazard (1982) claims that a comparable semantic effect to the one observed with locative arguments is at play here as well. Unlike their non-*rā*-marked counterparts, *rā*-marked temporal modifiers serve not only the temporal anchoring to the activity denoted by the predicate, but also its temporal delimitation: that is, the activity occupies the entire time interval. This contrast is illustrated by the difference of interpretation between examples (32d) and (33): (32d), but not (33), implies that the speaker has spent his life suffering.

(33) dar zendegi=am ranj bord-am
in life=CL.1SG pain carry.PST-1SG
'I have suffered in my life.'

In accordance with Lazard (1982), Ghomeshi (1997*b*) assumes that these modifiers behave as prototypical direct objects in that they "measure out" or delimit the event described by the verb (Ghomeshi & Massam 1994) and concludes that =*rā* is in fact a marker of high transitivity, since the cluster of the properties triggering its presence all correlate with high transitivity in the sense of Hopper & Thompson (1980).

3.3 *Rā* as a mark of topicality

The idea that DOM in Persian depends not only on the inherent referential features of the object but also on the information structure has been defended in several studies (Dabir-Moghaddam 1992, Dalrymple & Nikolaeva 2011, Ghomeshi 1997*b*, Karimi 1990, Lazard 1982, Meunier & Samvelian 1997, Peterson 1974, Shokouhi & Kipka 2003, Windfuhr 1979, among others). It has been observed that *rā*-marked objects tend to be topics, while non-*rā*-marked objects display focus properties. The set of data in ex. (34) unambiguously highlights the link between =*rā* and topicality. In these examples, =*rā* occurs with floating topics located at the left periphery of the sentence and cross-referenced by a clitic.

(34) a. u_i=rā gorosne va bi dud=esh_i
he=RA hungry and without smoke=CL.3SG
mi-gozāsht
IPFV-leave.PST.3SG

'He would leave him hungry and without opium.' (Lazard 1982: (83))

b. ān dokhtar$_i$=rā sad daf'e barā=yash$_i$ nāme
that girl=RA hundred time for=CL.3SG letter
nevesht-am
write.PST-1SG
'That girl, I have written one hundred letters to her.' (Lazard 1982: (87))

c. portegāl$_i$=o bāiad pust=esh$_i$=o kand ba'd khord
orange=RA must skin=CL.3SG=RA tear.SINF then eat.SINF
'As for oranges, one must peel them and then eat them.' (Lazard 1982: (92))

The floating topic can correspond to a number of different grammatical functions. In (34a), the topicalized constituent is the DO doubled by a clitic in the sentence, in (34b) the prepositional argument of the verb, and in (34c) the complement of the head noun of the direct object (i.e., it does not bear a function with respect to the verb).

Since the referential properties of an NP are generally correlated to its discourse status—definite/specific NPs are more likely to be topics while non-specific NPs tend to be focuses—some studies have claimed that the main function of =*rā* is to mark topicality. Peterson (1974) suggests that specific DOs are *rā*-marked because topics are specific in nature. Dabir-Moghaddam (1990, 1992, 2009), following Windfuhr (1979), claims that =*rā* is the mark of secondary topics: *all rā*-marked objects are topics while *all* non-marked objects are part of the comment. Dalrymple & Nikolaeva (2011) adopt a less "strong" version of this analysis. While they agree with Dabir-Moghaddam

(1992) that the main function of $=r\bar{a}$ is secondary topic marking, they nevertheless disagree with the latter on two points:

a) $=R\bar{a}$ can mark the primary topic as well.

b) While the distribution of $=r\bar{a}$ on non-objects exclusively depends on topicality, the picture is more complex for direct objects and topicality is not the only relevant factor in determining $r\bar{a}$-marking.

Following Reinhart (1982), Gundel (1988), and Lambrecht (1994), Dalrymple & Nikolaeva (2011) define topicality as a matter of "aboutness": the topic is the entity that the proposition is about. Consequently, topicality has to do with the construal of the referent as pragmatically salient (or prominent: de Swart 2007) so that the assertion is made about this referent. A potential diagnostic for topic-hood is the "what-about" or "as-for" test (Gundel 1988, Lambrecht 1994, Reinhart 1982). Although topicality correlates with the role played by the referent in the preceding discourse, the correlation is imperfect. Topicality is mainly a question of saliency, and although definite or specific NPs are generally given more saliency because of their referential properties, non-specific NPs may also become salient if the speaker decides so in a given communicative context.

The topic role is not necessarily unique and several studies have acknowledged the existence of at least a secondary topic along with the primary topic (Givón 1984, Nikolaeva 2001, Polinsky 1995). Nikolaeva (2001) defines secondary topic as "an entity such that the utterance is construed to be ABOUT the relationship between it and the primary topic". Topics are ordered with

respect to saliency: The primary topic is more pragmatically salient than the secondary topic.

Dalrymple & Nikolaeva (2011) note that =*rā* can be a primary topic as well as the secondary topic. In (35a), the *rā*-marked DO is a secondary topic, the subject being the primary topic, while in (35b) it is the primary topic given that the subject is the focus and therefore cannot be the primary topic.

(35) a. maryam ketāb=o be ki dād?
Maryam book=RA to who give.PST.3SG
'To whom did Maryam give the book?'

b. ki māshin=o did?
who car=RA see.PST.3SG
'Who saw the car?'

The second disagreement is more important. As mentioned above, Dalrymple & Nikolaeva (2011) claim topicality is not the only relevant factor in determining *rā*-marking on objects. It is a factor for some objects, i.e., indefinite objects. On definite objects, however, *rā*-marking is essentially motivated by definiteness, having to do with features of topic-worthiness: *all* definite objects must be marked, independent of their information status and even if they are in focus. In fact, Dalrymple & Nikolaeva (2011: 113) postulate two =*rā*'s or two functions for =*rā* in examples such as (36): the first marks the topicality of the temporal adjunct, while the second is licensed by definiteness.

(36) faghat in ye sā'at=o in ketāb=o be-khun
just this one hour=RA this book=RA IMP-read
'Read this book this one hour!'

Samvelian (2002) also defends the assumption that there are two =*rā*'s in Persian, noting a crucial difference between *rā*-marked arguments on the one hand and *rā*-marked floating topics and adjuncts on the other hand. While there can only be one *rā*-marked argument (i.e., object) in a simplex sentence, the number of *rā*-marked floating topics and adjuncts is not grammatically limited. The examples in (37) show that the nominal element of a complex predicate can be *rā*-marked if referential. The examples in (38) illustrate a "transitive" use of the same predicate, so that the verb is preceded by two direct nominal dependents. The relevant point here is that although the nominal element of the predicate can still be modified and even determined, ex. (38b), it cannot be *rā*-marked, ex. (38c). The situation is identical in "double object" constructions in (39). The complex predicate *pādāsh dādan* 'to reward' (lit. 'reward give') may take two direct objects (Theme and Goal), but only one of them can be *rā*-marked. When both arguments are definite, the dative (Goal) argument must have a prepositional realization.

(37) a. maryam ārāyesh karde bud
Maryam makeup do.PFP be.PST.3SG
'Maryam wore makeup.'

b. maryam zibā-tarin ārāyesh=rā karde bud
Maryam beautiful-SPR makeup=RA do.PFP be.PST.3SG
'Maryam wore the most beautiful makeup.'

(38) a. maryam in arus=rā ārāyesh karde bud
Maryam this bride=RA makeup do.PFP be.PST.3SG
'Maryam had made this bride up.'

b. maryam in arus=rā ārāyesh=e zibā=i karde
Maryam this bride=RA makeup=EZ beautiful=INDEF do.PFP

bud
be.PST.3SG
'Maryam made this bride up beautifully.'

c. *maryam in arus=rā zibā-tarin ārāyesh=rā karde
Maryam this bride=RA beautiful-SPR makeup=RA do.PFP
bud
be.PST.3SG
(intended) 'Maryam put the most beautiful make-up on this bride.'

or 'Maryam made this bride up in the most beautiful way.'

(39) a. maryam=rā ketāb=i pādāsh dād-and
Maryam=RA book=INDEF reward give.PST.3PL
'They rewarded Maryam with a book.' or 'They gave Maryam a book as a reward.'

b. be maryam in ketāb=rā pādāsh dād-and
to Maryam this book=RA reward give.PST.3PL
'They gave this book to Maryam as a reward.'

c. *maryam=rā in ketāb=rā pādāsh dād-and
Maryam=RA this book=RA reward give.PST.3PL

By contrast, *rā*-marked floating topics and adjuncts can be multiple: ex. (40) contains two *rā*-marked temporal modifiers and one *rā*-marked object.

(40) sāl=e pish=o, ruz-ā=ye jom'a=ro, bachch-hā-ro sinamā
year=EZ before=RA, day-PL=EZ Friday=RA, kid-PL=RA theatre
mi-bord-am
IPFV-take.PST-3PL
'Last year, on Friday evenings, I used to take the kids to the movies.'

To sum up, topicality cannot be considered to be the unique trigger of =*rā*, as claimed by Dabir-Moghaddam (1990, 1992, 2009). In fact, it would be nothing less than surprising if an item whose presence is mandatory in some cases was exclusively triggered by information structure. Information

structure is a matter of choice: the speaker decides how to "structure" the utterance in order to convey information. The presence of =*rā* with definite objects is obligatory in Persian. So, whatever the choice of the speaker about the information or discourse status of a definite object, =*rā* must be there as a grammatical constraint. The price for the analysis of =*rā* as an exclusive topic-marker would be to admit that Persian speakers have no choice as to the discourse status of the definite direct object, which must be necessarily construed as a topic. If definite subjects can be non-topical in Persian, why would definite objects be denied this status? As has been noted by Karimi (1989, 1990), a definite DO can be a focus. In (41), the *rā*-marked object can be the answer to a *who*-question and hence is the focus of the utterance. Consequently, topicality is not a necessary condition for *rā*-marking.

(41) diruz, tu kuche Maryam=o did-am
Yesterday in street Maryam=RA see.PST-1.SG
'Yesterday I saw Maryam in the street.'

There remains one last problem with the analysis of =*rā* as a topic-marker. Karimi (1989) rightly claims that not all non-subject topics are *rā*-marked, as illustrated by (42), where the bare object *gusht* 'meat' is topicalized without being *rā*-marked.

(42) gusht mi-dun-am ke maryam hichvaght
meat IPFV-know.PRS-1.SG that Maryam never
ne-mi-khor-e
NEG-IPFV-eat.PRS-3.SG
(lit.) 'Meat, I know Maryam never eats.'

More investigation is needed on the topicalization of 'bare' objects in

Persian, but one can already affirm that not only is topicality not a necessary condition for *rā*-marking, it is not either a sufficient condition.

3.4 Dom, object positions and word order

Several studies have noted that *rā*-marking has a syntactic correlate: *rā*-marked objects tend to precede prepositional arguments and even subjects and thus do not occur in the canonical position of DOs in Persian, i.e., adjacent to the verb. Based on this fact and some others that will be discussed later, many studies have suggested a dual-position account of the direct object depending on its markedness (Browning & Karimi 1994, Ganjavi 2007, 2011, Ghomeshi 1996, 1997*b*, Karimi 1990, 2003).[14] It is assumed that *rā*-marked objects do not occupy the same syntactic position as their non-*rā*-marked counterparts and appear in a higher position than the former. According to Karimi (2003), for instance, *rā*-marked DOs occupy the position of the specifier of VP, while non-*rā*-marked DOs occupy a lower position, that is, the position sister to the verb (under the V′):[15]

(43) a. [VP DP[$_{+Specific}$] [V′ PP V]]

b. [VP [V′ PP [V′ DP[$_{-Specific}$] V]]]

[14]Studies of phrasal syntax within the generative paradigm generally assume two distinct positions for objects, VP-internal and VP-external, and establish a correlation between object position and object marking (Diesing 1992, van Geenhoven 1998, Ritter & Rosen 2001, among many others): indefinite/nonspecific objects are generally assumed to be VP-internal, while marked objects are VP-external.

[15]In a more recent work, Karimi (2005) proposes a revised version of her Two Object Position Hypothesis (TOPH), in which both objects are base-generated in the same position, under the v'. The specific object shifts into the specifier of vP position in order to receive its interpretation.

The Two Object Position Hypothesis (TOPH) is built on the claim that *rā*-marked and non-*rā*-marked objects display consistent asymmetries with respect to word order, licensing parasitic gaps, and binding anaphors, and that they cannot be coordinated.

The unmarked word order asymmetry is the backbone argument of the TOPH. It is generally assumed that in unmarked, canonical or neutral word order in ditransitive constructions in Persian, *rā*-marked DOs precede the indirect object (IO) while non-*rā*-marked DOs follow the IO (Browning & Karimi 1994, Ganjavi 2007, Givi Ahmadi & Hassan 1995, Ghomeshi 1997*b*, Karimi 2003, Mahootian 1997, Rasekhmahand 2004, Roberts et al. 2009).

(44) a. kimeā aghlab barā mā she'r mi-khun-e
Kimea often for us poem IPFV-read-3SG
'It is often the case that Kimea reads poetry for us.'

b. kimeā aghlab barā mā ye she'r az Hafez mi-khun-e
Kimia often for us a poem of Hafez IPFV-read-3SG
'It is often the case that Kimea reads a poem by Hafez for us.'

(Adapted from Karimi 2003: 91–92)

(45) a. kimeā aghlab hame=ye she'r-hā=ye tāze-ash=ro barā mā mi-khun-e
Kimea often all=EZ poem-PL=EZ new=CL.3SG=RA for us IPFV-read-3SG
'It is often the case that Kimea reads all her new poems for us.'

b. Kimeā aghlab ye she'r az Hafez=ro barā mā mi-khun-e
Kimea often a poem of Hafez=RA for us IPFV-read-3SG
'It is often the case that Kimea reads a (particular) poem by Hafez for us.' (Adapted from Karimi 2003: 91–92)

Samvelian (2001) questions this hypothesis and postulates a flat struc-

ture for the Persian VP:[16] *rā*-marked and non-*rā*-marked objects occupy the same syntactic positions. In a series of recent corpus-based and experimental studies, Faghiri et al. (2014), and Faghiri & Samvelian (2014) show that, with respect to word order, indefinite non-marked DOs group with marked DOs rather than with bare objects. Faghiri & Samvelian (2015*b*) and Faghiri (2016) further argue that the asymmetries between marked and non-marked objects do not need to be accounted for in terms of syntactic positions and are best accounted for by semantic and discourse considerations.

Based on a corpus-based study[17] and experimental follow-up studies, Faghiri & Samvelian (2014), Faghiri et al. (2014) and Faghiri (2016) investigate ordering preferences between the DO and the IO in the preverbal domain in Persian. Observing that non-*rā*-marked objects show more versatility with respect to word order, the authors resort to a more fine-tuned classification of unmarked DOs, splitting them into bare DOs, ex. *ketāb* 'book', bare modified DOs, ex. *ketāb=e kohne* 'old book', and indefinite (unmarked) DOs *yek ketāb(=e kohne)* or *ketāb(=e kohne)=i* 'an (old) book'. In addition to the realization of the DO, these studies take into account other potentially influential factors such as relative length, givenness, collocationality, and lexical bias, via mixed-effect regression modeling, in line with key empirical studies on word order variations (Wasow 2002, among others).

The data reveal that while *rā*-marked DOs show a strong preference for

[16]Bonami & Samvelian (2015) also adopt a flat structure for Persian sentences.

[17]The study is based on the Bijankhan corpus, a corpus collected from daily news and common texts, in particular the newspaper *Hamshahri*, of about 2.6 million tokens, manually tagged for part-of-speech information. The corpus was created in 2005 by the DataBase Research Group at the University of Tehran and can be freely downloaded from their website.

appearing before the IO, among various non-*rā*-marked DOs, only bare nouns show a strong preference for adjacency to the verb. Interestingly, indefinite (non-*rā*-marked) DOs show a clear preference for the inverse, grouping with *rā*-marked DOs. Moreover, extra syntactic factors such as relative length also play a significant role in these ordering preferences. Accordingly, Faghiri & Samvelian (2015*b*,*a*) argue that the ordering preferences observed for different types of DOs are best represented as a continuum based on the degree of conceptual and/or discourse accessibility. The authors conclude that any structural account of word order preferences between DOs would lead to wrong predictions. Moreover, even if the TOPH was to be maintained, word order preferences speak in favor of an identical position for *rā*-marked DOs and indefinite non-*rā*-marked DOs. To sum up, word order does not seem to constitute a conclusive criterion in favor of a configurational account of *rā*-marking.

The behavior of DOs with respect to licensing parasitic gaps is another argument in favor of the TOPH. According to Karimi (1999: 704), only *rā*-marked DOs can license parasitic gaps, ex. (46).

(46) a. Kimea [$_{DP}$ in ketāb=ro]$_i$ [$_{CP}$ ghabl-az inke *pro* e$_i$
Kimea this book=RA before that
be-khun-e] be man dād
SBJV-read.PRS-3SG to I give.PST.3SG
'Kimea gave me this book before reading (it).'

b. * Kimea [$_{DP}$ ketāb]$_i$ [$_{CP}$ ghabl-az inke *pro* e$_i$
Kimea book before that
be-khun-e] be man dād
SBJV-read.PRS-3SG to I give.PST.3SG

Faghiri & Samvelian (2015*a*) note, however, that examples in (47) are

grammatical despite the fact that the non-*rā*-marked object licenses a parasitic gap. The oddness of (46b) may be due to the fact that the verb is in the past tense and the sentence denotes a specific accomplished event where it is expected for the DO to be known to the speaker and hence a bare DO is not felicitous.

(47) a. man [hendune]$_i$ [ghabl-az inke *pro* e$_i$ mazze kon-am]
I watermelon before that taste buy.PRS-1SG
ne-mi-khar-am
NEG-IPFV-buy.PRS-1SG
'I wouldn't buy watermelon(s) before tasting it/them.'

b. man [ye lebās]$_i$ [bedun=e inke *pro* e$_i$ emtehān kon-am]
I a cloth without=EZ that try do-PRS.1SG
kharid-am va hichvaght ham na-pushid-am=esh
buy.PST-1SG and never too NEG-wear.PST-1SG=CL.3SG
'I only bought a piece of clothing without trying (it) and I never wore it.'

Unlike non-*rā*-marked DOs, *rā*-marked DOs have been claimed to be able to bind an anaphor (Karimi 2003: 102):

(48) a. man [se-tā bachche-hā=ro]$_i$ [be hamdige]$_i$ mo'arrefi
I three-CLF child-PL=RA to each-other introduce
kard-am
do.PST-1SG
'I introduced the three children to each other.'

b. * man [se-tā bachche-hā]$_i$ [be hamdige]$_i$ mo'arrefi
I three-CLF child-PL to each-other introduce
kardam
do.PST-1SG

Here again, Faghiri & Samvelian (2015*a*) and Faghiri (2016) claim that in a proper context, a nonspecific DO can bind the IO, as shown by the attested

examples in (49), found on the Web.

(49) a. [chand varagh kāghaz]$_i$ [be hamdige]$_i$ mangane
some sheet paper to each-other staple
mi-kon-am
IPFV-do.PRS-1SG
(lit.) 'I staple a few sheets of papers to each other.'

b. lidya yeki=ro mi-shnās-e ke [dokhtar pesar]$_i$
Lidya someone=RA IPFV-know.PRS-3SG that girl boy
[be ham]$_i$ mo'arrefi mi-kon-e
to each-other introduction IPFV-do.PRS-3SG
'Lidya knows someone who introduces girls and boys to each other.'

The fact that non-*rā*-marked DOs and *rā*-marked DOs cannot be coordinated has also been evoked in favor of the TOPH (Karimi 2003: 103):

(50) *man diruz in aks=ro va ketāb kharid-am
I yesterday this picture=RA and book buy.PST-1SG

Samvelian (2001) claims that coordination cannot be used as a test in favor of the TOPH, and Faghiri & Samvelian (2015*a*) give the following grammatical example in which a non-*rā*-marked and a *rā*-marked DO are coordinated:

(51) man diruz sham'=o ye rumizi=o in tāblo=ro
I yesterday candle=and a tablecloth=and this painting=RA
kharid-am
buy.PST-1SG
'Yesterday, I bought this painting, a tablecloth, and some candles.'

To sum up, although there is a consensus on the TOPH in the studies within the generative framework, at least some of the empirical facts supporting this hypothesis seem to be fragile. In particular, word order preferences

do not allow for a clear-cut distinction between *rā*-marked DOs and their non-*rā*-marked counterparts as far as syntactic positions are concerned.

3.5 Concluding remarks on DOM

The facts addressed in this section show that despite the abundant literature on the semantic and pragmatic parameters triggering =*rā*, there is still a lot to investigate in order to draw a clear picture of the situation. Therefore, it is a safe bet that =*rā* will remain a popular issue in forthcoming studies on Persian. However, it seems reasonable to conclude, in line with Lazard (1982) and Ghomeshi (1997*b*), that rather than there being a single binary feature that can characterize *rā*-marking, be it specificity, topicality, any other feature, the presence of =*rā* is determined by the interaction between several parameters that have been highlighted in various studies.

Complex as it may seem, this situation is neither specific to Persian nor to DOM. Bossong (1991) concludes that in many languages the rules of DOM cannot be formulated precisely, but must allow for a certain degree of variability across speakers and situations. Variability is also observed for phenomena other than DOM, like word order, optional realization of some appositions, dative alternation, etc. A growing body of studies since Wasow (2002), Bresnan (2006), Bresnan et al. (2007), and Bresnan & Hay (2008), among others, accounts for grammatical phenomena that involve variation by resorting to a new approach which assumes that variation is part of grammar and can be statistically modeled. These methods can be very useful for the study of *rā*-marking in Persian, which involves various parameters whose complex

interaction requires more reliable methods of investigation than traditional grammaticality judgments. This point has been clearly demonstrated by Faghiri & Samvelian (2014, 2015*b*,*a*), Faghiri et al. (2014), and Faghiri (2016) for the linear position of the *rā*-marked and non-*rā*-marked objects: generalizations based on grammaticality judgements turn out to be (partially) wrong when rigorous empirical methods are used. The same vein of research can be applied for modeling semantic and discourse dimensions of *rā*-marking.

4 Complex predicates

Persian has only around 250 simplex verbs, half of which are currently used by the speech community[18]. The morphological lexeme formation process outputting verbs from nouns (ex. *khāb* 'sleep' > *khāb-idan* 'to sleep', *raghs* 'dance' > *raghs-idan* 'to dance'), though available, is not productive. When they need to refer to a new event type, speakers resort to complex predicates (CPrs), formed by a verb and a preverbal element, which can be a noun, ex. *harf zadan* 'to talk' (lit. 'talk hit'), an adjective, ex. *bāz kardan* 'to open' (lit. 'open do'), a particle, ex. *bar dāshtan* 'to take' (lit. 'PARTICLE have'), or a prepositional phrase, ex. *be kār bordan* 'to use' (lit. 'to work take'). These combinations are generally referred to as complex predicates, compound verbs, or light verb constructions.

According to Telegdi (1951), the gradual elimination of simplex verbs and their substitution by "periphrastic expressions" or "compound verbs" is

[18]Sadeghi (1993) gives the estimate of 252 verbs, 115 of which are commonly used. Khānlari (1986) provides a list of 279 simplex verbs. The Bijankhan corpus contains 228 lemmas.

at least as old as Middle Persian. Korn (2013) argues that the rise of CPrs in Persian is linked to the development of the verb pair "do" and "become", which encode the features called Instigation [+INST] and Affectedness [+AFF], respectively.

A first issue when dealing with Persian CPrs is the delimitation of the category itself. As discussed extensively in Samvelian (2001, 2012), two facts drastically blur the boundary line between lexical and light verbs, and hence, between CPrs and ordinary object-verb combinations:

1. The (expected) consequence of the limited number of simplex verbs in Persian is that most of them have a vague semantic content, which becomes specified only in the context of their combination with their arguments (Samsam Bakhtiari 2000, Samvelian 2012). In other words, *most* Persian verbs are *de facto* light verbs, so that, from a semantic point of view, deciding whether a noun-verb combination qualifies for being a CPr involves some degree of arbitrariness. For instance, in combinations like *rang zadan* 'to paint' (lit. 'color hit'), *vāks zadan* 'to polish' (lit. 'polish hit'), or *kare zadan* 'to butter' (lit. 'butter hit'), the verb *zadan* may be considered either a "bleached" (light) verb or a lexical verb meaning 'to apply, to put', and accordingly the sequence can be considered as a CPr or an ordinary object-verb combination. The most striking piece of evidence illustrating this situation is the variability of the combinations listed in Persian dictionaries, which vary considerably from one dictionary to another.

2. From a strictly syntactic point of view, [bare object - lexical verb] com-

binations and [adjective - copula] combinations are in many respects comparable to N-V and A-V CPrs. For instance, a part of the criteria used to identify N-V CPrs, like single word stress and limited syntactic autonomy for the noun, also apply to [bare object - lexical verb] combinations, leading to a situation where sequences like *māshin rāndan* 'to drive a car' are considered 'compound verbs' in some dictionaries and in the literature on CPrs. Dabir-Moghaddam (1995), for instance, suggests that sequences like *ruznāme xāndan* 'to read a newspaper' are also compound verbs comparable to sequences like *ersāl kardan* 'to send' (lit. 'sending do').

Identifying the type of constructions that can be considered as CPrs has thus been one of the issues discussed in some studies (Dabir-Moghaddam 1995, Samvelian 2012, Sedighi 2009, *among others*). However, this issue is probably far from being resolved, probably because whether a sequence is a CPr or not – except for uncontroversially clear combinations formed with a "real" light verb such as *kardan* and a "predicative" (or eventive) noun, such as *ersāl* 'sending' – is often a matter of usage and lexicalization rather than the inherent properties of the sequence itself.

The main body of work on CPrs has focused on the dual nature of these sequences, which exhibit both lexical and phrasal properties (Barjasteh 1983, Dabir-Moghaddam 1995, 1997, Family 2006, 2009, Folli et al. 2005, Goldberg 1996, 2003, Karimi 1997, Karimi-Doostan 1997, 2005, Lazard 2013, Megerdoomian 2001, 2002, 2012, Müller 2010, Pantcheva 2010, Samsam Bakhtiari 2000, Samvelian 2012, Samvelian & Faghiri 2013*b*, 2014, Shabani-

Jadidi 2015, Tabaian 1979, Vahedi-Langrudi 1996, among others). On the one hand, Persian CPrs display all properties of syntactic combinations, including some degree of semantic compositionality, while, on the other hand, they also have word-like properties, since CPr formation has all the hallmarks of a lexeme formation process, such as lexicalization, idiomaticity, and the fact that the sequence can undergo morphological operations.

4.1 Words or phrases?

Whether Persian CPrs are "words" or syntactic construals has been one of the most debated issues in the literature. The following arguments are generally put forward in favor of a lexical view of these sequences:

- The whole sequence bears a single lexical stress, like a word, ex. (52).
- CPrs can serve as input to derivational rules, ex. (53).
- The components of a CPr must be adjacent and can only be separated by a restricted set of elements, i.e., verbal inflectional prefixes, clitic pronouns and the future auxiliary, ex. (54).[19]

(52) and`ākht — throw.PST.3SG — 'threw'

b. be gery`e andākht — to cry throw.R2 — 'made cry' (Ghomeshi & Massam 1994)

(53) Adjective derivation with the suffix -*i*

a. khordan 'to eat' >

khordan-i 'edible'

[19]This claim is questioned in several studies and will be discussed later.

b. dust dāshtan ‘to love’ (lit. ‘friend have’) >

dust dāshtan-i ‘lovely’

(54) a. maryam kheili gol dust dār-ad
Maryam much flower friend have.PRS-3SG

b. # maryam gol dust kheili dār-ad
maryam flower friend much have.PRS-3SG
‘Maryam like flowers very much.’

(55) a. maryam gol dust na-dār-ad
Maryam flower friend NEG-have.PRS-3SG
‘Maryam doesn’t like flowers.’

b. maryam dust=ash dār-ad
Maryam friend=CL.3SG have.PRS-3SG
‘Maryam loves her/him/it.’

c. maryam gol dust khāh-ad dāsht
Maryam flower friend FUT.AUX-3SG have.SINF
‘Maryam will love flowers.’

Arguments in favor of a phrasal (i.e., syntactically construed) analysis of CPrs are the following:

a. Not only inflectional material, but also syntactic material such as prepositional phrases, can intervene between the verb and the non-verbal element in a CPr, ex. (56).

b. The non-verbal element, if a noun, can be modified, quantified, determined, and *rā*-marked, ex. (57).

(56) a. hichvaght be ātish dast na-zan
never to fire hand NEG.IMP-hit.PRS.2SG

b. hichvaght dast be ātish na-zan
never hand to fire NEG.IMP-hit.PRS.2SG
'Never touch the fire.'

(57) a. maryam harf zad
Maryam speech hit.PST.3SG
'Maryam talked.'

b. maryam in harf-hā=rā zad
Maryam this speech-PL=RA hit.PST.3SG
'Maryam told these words.'

These apparently contradictory properties have given rise to debates on the appropriate "place" for CPr formation. "Lexicalist" approaches[20] claim that Persian CPrs are formed in the lexicon (Barjasteh 1983, Dabir-Moghaddam 1995, 1997, Karimi-Doostan 1997). Karimi-Doostan (1997, p. 193) suggests that Persian CPrs are lexically formed complex lexical entries which consists of two zero-level elements separable in syntax, and thus failing to display lexical integrity. Goldberg (1996) treats the Persian CPr as a construction represented in the lexicon, whose categorial status is V° by default. This guarantees that the verb and the preverbal element are unseparated and thus can undergo derivational processes. However, the V° status is a default status and can be overridden if there is a competing higher-ranked constraint.

"Syntactic" approaches, by contrast, consider CPr formation a syntactic process (Folli et al. 2005, Ghomeshi & Massam 1994, Megerdoomian 2002, 2012, Mohammad & Karimi 1992, Tabaian 1979, Vahedi-Langrudi 1996,

[20]The term "lexicalist" is ambiguous in the literature. Here, it must be understood as "morphologically formed".

among others). Mohammad & Karimi (1992) suggest that despite their syntactic construal, Persian CPrs display lexical properties for two reasons: 1) the impossibility for the light verb to assign thematic roles, and 2) the existence in Persian of two distinct positions for objects.

More recently, neo-constructionist studies on Persian CPrs (Folli et al. 2005, Megerdoomian 2002, 2012, Pantcheva 2010) adopt the predicate decomposition approach developed by Hale & Keyser (1992, 1993, 1997, 2002), Ritter & Rosen (1996) and Borer (1994), which blurs the classic distinction between simplex and complex predicates and analyzes pairs like *give a kick* and *kick* in English as sharing the same syntactic representation: the "simplex" verb *kick* is syntactically construed via the incorporation of the noun *kick* into an abstract light verb. Megerdoomian (2002) and Folli et al. (2005) claim that Persian CPrs constitute a conclusive argument in favor of neo-constructionist theories of argument structure, since they are unincorporated counterparts of English simplex verbs and thus reveal the universally complex underlying structure of predicates, be they morphologically simplex or not.

On this view, whether CPrs are formed in the lexicon (i.e., morphologically) or in syntax is not relevant, since morphology is handled in syntax.

Samvelian (2012) also considers the debate on the dual nature of Persian CPrs a false issue, not because of the lack of a boundary between syntax and morphology, but because of the numerous confusions surrounding the use of terms such as "formed in the lexicon," "word", and "morphologically formed." Saying that a sequence containing more than one word is "formed

in the lexicon" or is a "lexical unit" can mean various things:

a. The sequence is the output of a morphological operation and prototypically behaves like an atom in syntax.

b. The sequence is lexicalized (Aronoff 1993) for various reasons (token frequency, naming force, etc.) and must be stored. It is a listeme in the sense of Di Sciullo & Williams (1987).

These are two independent dimensions and must be carefully distinguished. Gaeta & Ricca (2009, p. 38) suggest a quadripartite typology, which allows for treating the properties of being a lexical/stored unit or the output of a morphological operation as independent grades of freedom. Each dimension is represented by a binary feature, [± morphological] and [± lexical], giving rise to four virtually possible combinations. This typology shows that the set of morphological words and the set of lexicalized items need not be coextensive.

Turning now to Persian CPrs, the latter are certainly not words in sense (a), since even the most idiomatic Persian CPrs do not behave as atoms and are separable by lexical material. Most Persian CPrs, on the other hand, are lexicalized sequences and display lexeme-like (i.e., not word-like) properties (Bonami & Samvelian 2010, Samvelian & Faghiri 2013*a*). They are [− morphological], [+ lexical] in the sense of Gaeta & Ricca (2009). Samvelian (2012) furthermore argues that none of the arguments supporting the "wordhood" of Persian CPrs are conclusive:

- **Lexical accent**. Bearing a single lexical stress is not specific to CPrs.

Sequences formed by a bare object and a lexical verb bear a single stress and yet have not claimed to be "words".[21]

- **Input to morphological operations**. It has been argued that since Persian CPrs can undergo morphological operations, such as nominalization, "they must be treated as lexical or X° units" (Megerdoomian 2002: 59). For Karimi-Doostan (1997), given that the agentive noun °*konande* 'doer' is not attested, the agentive noun *pazira'i konande* 'entertainer' must be derived from *pazira'i kon*, which requires that the latter be a word. According to Vahedi-Langrudi (1996) the suffix -*i* adjoins to the whole sequence *eslāh kardan* 'to reform', and not to *kardan* in order to form *eslāh kardani* 'likely to be reformed'.

  Samvelian (2006*a*, p. 162) argues, however, that this line of argumentation is flawed since it leads to the conclusion that *mār zadan* 'snake beat' is a lexical unit on the basis of the existence of the adjectival participle *mār zade* 'snake beaten'. But *mār zadan* never occurs as a sequence in discourse.

  Bracketing paradoxes, i.e., cases where the semantic scope of an affix and its morphological attachment do not coincide (Pesetsky 1985, Spencer 1988, Sproat 1984, Williams 1981), are rather common in various languages. In the French agentive noun *metteur en scène* 'director, producer', the derivational affix -*eur* is attached to the verbal stem, while its scope is the whole sequence. Note that the agentive noun °*metteur* 'putter' (from the verb *mettre* 'to put') is not attested, like

[21] Ghomeshi & Massam (1994) mention also this fact in favor of a syntactic analysis of Persian CPrs.

°*konande* 'doer' in Persian. Yet, it has not been suggested to derive *metteur en scène* from *mettre en scène*. Such an analysis would imply that the affix -*eur* behaves like an infix and interrupts a 'word'. Therefore, the only option here is to derive *metteur* morphologically first. A similar analysis has been outlined for producing sequences such as *pazira'i kon + -ande* 'entertainer' by Müller (2010) within the HPSG framework. A lexical rule applies to the stem of *kardan* 'to do' first and produces *konande* 'doer'. Since the lexical entry for *kardan* specifies that it must combine with a preverbal element to form a CPr, *konande* inherits this information and combines in turn with *pazira'i*.

- **Inseparability**. Several studies affirm that the components of a CPr can only be separated by a restricted set of items, which are either morphological material (affixes) or grammaticalized lexical material, i.e., the future auxiliary, comparable to inflectional affixes (Dabir-Moghaddam 1995, Goldberg 1996, 2003, Karimi-Doostan 1997). The insertion of "real" syntactic items, claim these studies, is excluded and gives rise to ungrammatical or odd examples, ex. (58b).

(58) a. Ali=rā setāyesh kard-am
Ali=RA adoration do.PST-1SG
'I adored Ali.'

b. ?? Setāyesh Ali=rā kard-am
adoration Ali=RA do.PST-1SG
(Goldberg 1996: 135)

Other studies, however, question this claim and affirm that the members of a CPr can be separated by syntactic or lexical items (Sami-

ian 1983, Ghomeshi & Massam 1994, Ghomeshi 1996, Samvelian 2001, 2012). The following examples are from Ghomeshi (1996):

(59) a. gush aslan ne-mi-kon-e
ear absolutely NEG-IPFV-do.PRS-3SG
'(S)he never listens.'

b. garm=esh tu madrese mi-kon-am
warm=CL.3SG in school IPFV-do.PRS-1SG
'I'll heat it at school.'

Samvelian (2012) also provides numerous attested examples where the prepositional argument of the CPr interrupts the latter (Samvelian 2012: 58–60):

(60) a. ān zan bargasht va sili be surat=am
that woman turn.PST.3SG and slap to face=CL.1SG
zad
hit.PST.3SG
'The woman turned around and slapped me.' (S. Daneshhvar, *Savushun*)

b. sedā=ye āhan ra'she be tan=am
sound=EZ iron shiver to body=CL.1SG
mi-andāz-ad
IPFV-throw.PRS-3SG
'The sound of the iron makes me shiver.' (Sh. Mandanipour, *Shargh-e Banafshe*)

To conclude, none of the properties of Persian CPrs provide a conclusive argument in favor of their analysis as "words" or as being "morphologically" formed. Persian CPrs display all typical properties of syntactic combinations

and parallel object-verb combinations in Persian. However, although not words, Persian CPrs are clearly multiword expressions and CPr formation has all the trappings of a lexeme formation process. The lexical properties of CPrs result from their being lexemes or "phrasal lexemes" in the sense of Masini (2009).

4.2 The syntactic status of the nominal element

On the basis of a thorough comparison between the nominal element of the CPr and the bare object of a lexical verb, Samvelian (2012) shows that the former are syntactically comparable to bare objects in all respects. Like bare objects, the nominal element of a CPr:

1. is generally adjacent to the verb and tends to follow adverbials and prepositional arguments.
2. rarely (if not never) appears in the post-verbal position, ex. (61).
3. can be fronted (extracted) and receive a topical reading, ex. (62).
4. can be promoted and become the subject of the passive construction, ex. (63).
5. can be coordinated with the nominal element of another CPr, ex (64).

(61) ?? in lebās mi-dah-ad bu
this dress IPFV-give.PRS.3SG smell
(intended) 'This dress smells (bad).'

(62) dast man bar-hā be=het goft-am ke be ātish
hand I time-PL to=CL.2SG tell.PST-1SG that to fire
na-zan
NEG.IMP-hit.PRS.2SG
'I've told you several times not to touch the fire.'

(63) a. matbu'āt be in mas'ale kheili ahammiyat dād-and
press to this issue much importance give.PST-3.PL
'The press gave much importance to this issue.'

b. be in mas'ale kheili ahammiyat dāde shod
to this issue much importance give.PFP become.PST.3SG
'Much importance was given to this issue.'

(64) a. omid lagad va sili khord
Omid kick and slap collide.PST.3SG
'Omid was kicked and slapped.'

Samvelian (2012) concludes that the nominal element of the CPr has exactly the same syntactic status as a bare direct object.[22] The differences between the latter and the former are a matter of semantics and not of syntactic construal. While the noun in a CP is more cohesive with the verb than a bare direct object (in terms of word order, differential object marking, and pronominal affix placement), it is impossible to draw a categorical syntactic distinction between the two types of combinations.

4.3 Compositionality, productivity and idiomaticity

The compositionality of Persian CPrs has received a great deal of attention in recent literature. Although Persian CPrs are idiomatic, they are also highly

[22]For a detailed description of bare nouns in Persian, see also Modarresi (2014).

productive. Several studies have suggested that compositionality is the key to this productivity and suggested hypotheses on how the contribution of the verb and the preverbal element must be combined to derive the meaning, or at least some of the semantic properties, of the CPr.

Two main arguments have been invoked in favor of a compositional analysis of Persian CPrs:

a. The predictability of their argument and event structure.

b. The predictability of their lexical (referential) meaning.

In examples (65) and (66) below:

a. The referential meaning of the CPr and the roles assigned to the arguments are determined by the nominal element, since the semantic participants of the CPr *sili zadan* 'to slap' (Lit 'slap hit') in (65b) are identical to those realized within the NP headed by *sili* 'slap' in (65a).

b. The verb, on the other hand, determines the argument mapping, since the substitution of *zadan* 'to hit' in (65b) for *khordan* 'to collide' in (65c) entails a change in the mapping of the participants to grammatical functions.

c. The verb also seems to determine some of the aspectual properties of the CPr, since the verb alternation in (66), *dāshtan* 'to have' vs. *āvardan* 'to bring', gives rise to an aspectual contrast.

(65) a. sili=e sārā be omid
slap=EZ Sara to Omid
'Sara's slap to Omid'

b. sārā be omid sili zad
Sara to Omid slap hit.PST.3SG
'Sara slapped Omid.'

c. omid az sārā sili khord
Omid from Sara slap collide.PST.3SG
'Omid was slapped by Sara.' (Samvelian & Faghiri 2014: 45 (1))

(66) a. maryam (hamishe) in ettefāgh=rā be yād dāsht
Maryam always this event=DOM to memory have.PST.3SG
'Maryam (always) remembered this event.' (durative reading)

b. maryam (nāgahān) in ettefāgh=rā be yād āvard
Maryam suddenly this event=RA to memory bring-PST.3SG
'Maryam (suddenly) remembered this event.' (punctual reading)

(Samvelian & Faghiri 2014: 45 (1))

Various approaches have been developed to account for these facts. Most studies within the generative framework adopt a fully compositional view, in the sense that they build on the assumption that the respective contributions of the components of a CPr are consistent through all their combinations and can be defined *a priori*. Projectionist approaches, for example Karimi-Doostan (1997), assume that the information stored in the lexical entries of the light verb and the non-verbal element combine to build a CPr, while constructionist approaches, for example Megerdoomian (2001, 2002, 2012), Folli et al. (2005), and Pantcheva (2010), consider the syntactic and the semantic properties of a CPr to be derived from the syntactic construction in which the verb and the preverbal element are inserted.

Alternative analyses have been developed in studies adopting a construction-based approach in the sense of Fillmore et al. (1988), Goldberg (1995), and Kay & Fillmore (1999). These studies account for the productivity of Per-

sian CPrs either by adopting a non-compositional view (Family 2006, 2009, 2014) or by developing a different view of compositionality (Samvelian 2012, Samvelian & Faghiri 2013*a*,*b*, 2014).

Karimi-Doostan (1997) provides one of the first serious attempts to model the respective contributions of the verb and the non-verbal element in CPr formation. Based on Butt's (1995) work on argument structure, Karimi-Doostan proposes an account in terms of argument "fusion" or "reformation". Following Grimshaw & Mester (1988), he assumes that the light verb (LV) does not assign theta-roles and therefore does not have an argument structure. However, it displays aspectual properties and assigns an aspectual role. Being thematically defective, the LV must combine with another element, namely the preverbal element of the predicate, to develop into a syntactically and semantically complete verb. This combination gives rises to two kinds of CPrs, either compositional or non-compositional. The first kind results from the combination of the LV with a predicative noun, that is, a noun displaying an argument structure, such as *sili* 'slap' in (65). Non-compositional CPrs are formed when the LV combines with a "thematically opaque" noun, i.e., a noun that does not display an argument structure. *Yax zadan* 'to freeze' (lit. 'ice hit'), *ghofl kardan* 'to lock' (lit. 'lock do') and *āb dādan* 'to water' (lit. 'water give') are examples of non-compositional CPrs. CPr formation involves the fusion of the information encoded in the respective lexical entries of the verb and the noun.

LVs are divided into three categories with respect to their aspectual properties: Initiatory, ex. *dādan* 'to give'; Transition, ex. *khordan* 'to collide';

and Stative, ex. *dāshtan* 'to have'. Some verbs may belong to more than one category, *kardan* 'to do' for example, which is either Initiatory or Transition, and thus have two lexical entries. The aspectual category of the LV determines the aspectual type of the CPr. Initiatory verbs form CPrs with at least one external argument, i.e., either unergative or transitive CPrs, and are compatible with nouns having at least one external argument that refers to the initiator of the action denoted by the CPr.

Transition verbs form CPrs with a single internal argument, i.e., unaccusative predicates, and are compatible with nouns having at least one (internal) argument. The latter is mapped into the subject function and receives the Patient role. A mapping rule ensures the correct association between an LV and a preverbal element. For instance, a noun like *shekast* 'defeat', which assigns Agent and Patient thematic roles, can combine with either an Initiatory verb or a Transition verb. In the first case, its external argument (i.e., the Agent) is mapped into the subject function, ex. (67a), while in the second case, it is the internal argument that becomes the subject, ex. (67b):

(67) a. ali sāsān=rā shekast dād
Ali Sasan=RA defeat give.PST.3SG
'Ali defeated Sasan.'

b. sāsān az ali shekast khord
Sasan from Ali defeat collide.PST.3SG
'Sasan was defeated by Ali.'

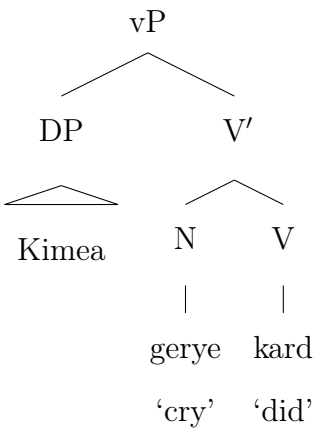
Since Karimi-Doostan (1997) postulates a categorical distinction between compositional and non-compositional CPrs, the latter are not accounted for by his treatment, even though he acknowledges that some of the regularities

mentioned above also hold for non-compositional CPrs.

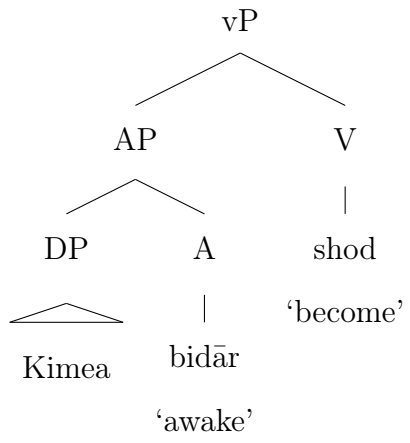
Megerdoomian (2001, 2012) and Folli et al. (2005) are representative examples of constructionist approaches to Persian CPrs. Based on work by Hale & Keyser (1993, 2002) and Borer (1994), these studies claim that the syntactic and the semantic properties of a CPr are derived from the syntactic construction in which the verb and the preverbal element are inserted, and not from their respective lexical entries. A fully compositional approach is thus maintained, but the burden shifts from the lexicon to the syntax. According to Folli et al. (2005), the verb in the CPr realizes the v head in Hale and Keyser's approach, as illustrated in (68)–(70).

Persian CPrs are thus the non-incorporated counterpart of verbal constructions suggested by Hale and Keyser.

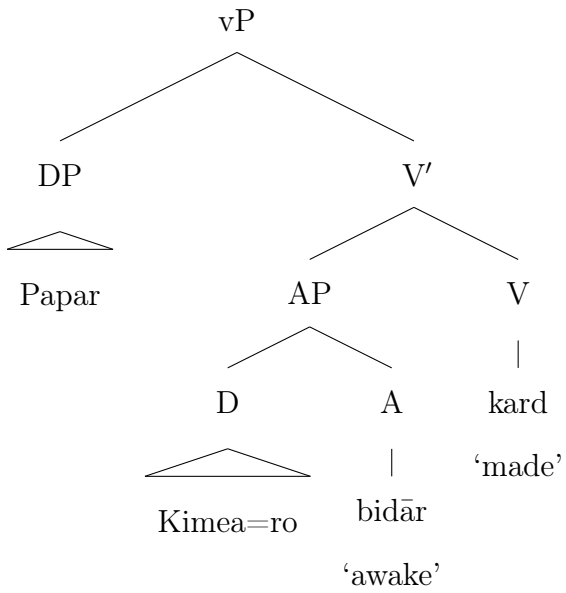
(68) Folli et al. (2005: 1374, (15b))



(69) Folli et al. (2005: 1375, (16b))



(70) Folli et al. (2005: 1375, (17b))



In this approach, the thematic role of Agent/Cause is assigned by v to its external argument (Kratzer 1996, Marantz 1997): *kardan* in (68) and (70) forms agentive predicates, but *shodan* in (69) does not . In other words, the LV, being the lexical realization of v, is responsible for the agentive properties of the CPr, while the non-verbal element plays no role here. Megerdoomian (2001: 69) argues along the same lines: "(...) the choice of the light verb

determines whether an external argument is projected". This claim is supported by the fact that changing the verb in a CPr entails a change in the mapping of the arguments to grammatical functions, as illustrated in (67).

Following Bashiri (1981), the authors claim that the verb also determines some aspectual properties of the CPr, namely its dynamic vs. stative and durative vs. punctual aspect. This explains the aspectual contrast between (67a) and (67b) above. Like Agent selection, aspectual/eventive properties of a given LV are assumed to be consistent through all its combinations to form a CPr. For instance, *dāshtan* 'to have' is always stative.

The non-verbal element, on the other hand, is claimed to determine the *Aktionsart* properties, i.e., telicity, and the referential meaning of the CPr. If the non-verbal element of the CPr is a PP, a particle, an adjective, or an eventive noun, the CPr is telic; otherwise—that is, if the nonverbal element is a non-eventive noun—the CPr is atelic.

Table 1, Adapted from Folli et al. (2005), resumes and exemplifies the contribution of each component in the makeup of the CPr.

Other constructionist analyses have been developed by Megerdoomian (2001, 2002, 2012) and Pantcheva (2010). Notwithstanding their differences, these approaches all build on the assumption that the respective contribution of the components participating in CPr formation is consistent through all their combinations and can be defined *a priori*.

In a series of studies, Samvelian (2012) and Samvelian & Faghiri (2013*a*,*b*, 2014) develop an alternative view of compositionality, which they qualify as *a posteriori* in the sense of Nunberg et al. (1994) for *idiomatically combin-*

Table 1: adapted from Folli et al. (2005: 1386, Table 3)

Telic Complex Predicates	
PP + LV:	*be donyā āmadan* 'to be born' (lit. 'world come') *be ātash keshidan* 'to put on fire (lit. 'fire pull')
Particle + LV:	*kenār āmadan* 'to get along' (lit. 'side come')
Adjective + LV:	*derāz keshidan* 'to lay down' (lit. 'long pull')
Eventive noun + LV:	*shekast khordan* 'to be defeated' (lit. 'defeat collide') *shekast dādan* 'to defeat' (lit. 'defeat give')
Atelic Complex Predicates	
Non-eventive noun + LV:	*dast khordan* 'to get touched' (lit. 'hand collide') *kotak khordan* 'to get beaten' (lit. 'beating hit') *dād zadan* 'to yell' (lit. 'scream hit') *dast andākhtan* 'to mock' (lit. 'hand throw')

ing expressions, and outline a construction-based analysis of Persian CPrs, building on the following observations:

1. Although there are consistent regularities in the makeup of the syntactic and semantic properties of CPrs, several examples show that the contribution of each component cannot be determined *a priori*, but is determined in combination with the other component of the CPr and the meaning of the construction as a whole.

2. While the idiomatic properties of Persian CPrs have been generally acknowledged, they have nevertheless been overlooked or minimized by studies that adopt a fully compositional approach.

Samvelian (2012), on the basis of extensive data, shows that the same verb can give rise to CPrs with different agentive and eventive properties. Likewise, the non-verbal element's contribution can vary through its combi-

nations with different verbs.

For instance, the verb *zadan* 'to hit', generally considered agentive and eventive, can nevertheless participate in the formation of "unaccusative" (or passive-like) CPrs such as *yax zadan* 'to freeze' (lit. 'ice hit') or *zang zadan* 'to go rusty' (lit. 'rust hit'). The same holds for *gereftan* 'to take' and *kardan* 'to do', which, apart from agentive CPrs, form "unaccusative" CPrs also, ex. *ātash gereftan* 'to take fire' (lit. 'fire take'), *ādat kardan* 'to get used to' (lit. 'habit do') and *dard kardan* 'to ache' (lit. 'pain do'). The verbal contribution is not consistent either with respect to the eventive properties of the CPr. Again, the same verb can give rise to both stative and eventive (dynamic) CPrs. For instance, the verb *dāshtan* 'to have' is not invariably stative and can produce eventive predicates, such as *ersāl dāshtan* 'to send' (lit. 'sending have'), *taghdim dāshtan* 'to offer' (lit. 'offering have'), and *e'lām dāshtan* 'to announce' (lit. 'announcing have').[23]

The contribution of the non-verbal element also turns out to be inconsistent. For instance, adjectives and PPs can as well form atelic CPrs, ex. *lāzem dāshtan* 'to need' (lit. 'necessary have'), *penhān dāshtan* 'to keep hidden' (lit. 'hidden have'), and *be maskhare gereftan* 'to make fun of' (lit. 'to mockery take'). Inversely, non-eventive nouns can give rise to telic CPs, *pust andākhtan* 'to slough off' (lit. 'skin throw').

The non-predictability of the meaning of the CPr is another significant impediment to fully compositional approaches. In order for the latter to

[23]Note that the examples discussed in this section are by no means isolated. For thorough examples illustrating the non-consistency of the verbal contribution to the agentive and eventive properties of Persian CPrs, see Samvelian (2012: 114–130).

work, the meaning of the CPr must be derivable on the basis of the meaning of its components. However, as mentioned in several studies (Bonami & Samvelian 2010, Family 2006, 2009, 2014, Goldberg 1996, Karimi-Doostan 1997, Samvelian 2012, Samvelian & Faghiri 2013*a*, among others), numerous Persian CPrs are semantically opaque. Moreover, as shown by Samvelian (2012) and Bonami & Samvelian (2010), even in cases where a CPr is semantically transparent, it is hardly ever the case that its meaning is fully predictable from the meaning of its component parts. In other words, the meaning of Persian CPrs, even the transparent ones, is conventional in many cases and therefore has to be learned, in the same way as one has to learn the meaning of the simplex verbs in English, for instance.

Relying on these observations, Samvelian (2012) and Samvelian & Faghiri (2013*a*,*b*, 2014) claim that Persian CPrs, at least the lexicalized ones, must be stored, exactly as lexemes are. They nevertheless argue that the need for an inventory is not incompatible with a compositional approach, provided compositionality is defined *a posteriori*. With respect to their compositionality, Persian CPrs are comparable to Idiomatically Combining Expressions, that is, "idioms whose parts carry identifiable parts of their idiomatic meanings" (Nunberg et al. 1994: 496). This means that the verb and the non-verbal element of a CPr can be assigned a meaning in the context of their combination. Thus, the CPr is compositional, in the sense that the meaning of the CPr can be distributed to its components, and yet it is idiomatic, in the sense that the contribution of each member cannot be determined out of the context of its combination with the other one. For instance, *zadan* 'to hit' can receive various interpretations according to the noun with which it com-

bines: 'to apply' in *rang zadan* 'to paint'; 'to add, to incorporate' in *namak zadan* 'to salt'; 'to wear' in *māsk zadan* 'to wear a mask'; or 'to emit' in *dād zadan* 'to shout'. Given the meaning assigned to *zadan* and the meaning of the CPr as a whole, new combinations can be produced and interpreted. For instance, *tag zadan* 'to tag' (lit. 'tag hit'), formed with the loanword *tag*, is created on the basis of *barchasb zadan* 'to label' (lit. 'label hit'), *tambr zadan* 'to stamp' (lit. 'stamp hit'), etc.

This view of Persian CPrs is then developed into a Construction-based approach:

- Each CPr corresponds to a Construction.
- CPrs can be grouped into classes according to their semantic and syntactic properties and each class can be represented by a partially fixed Construction.
- Constructions can be structured in networks, thus accounting for different semantic and syntactic relations between CPrs, such as synonymy, hyperonymy/hyponymy, and valency alternation.[24]

In this approach, the productivity of the Persian CPrs is accounted for via the analogical extension of the existing classes. It can be compositionality-based or not. In the first case, new combinations are created on the basis of the meaning assigned to the Construction as a whole and to its components. However, productivity is not always compositionality-based, and

[24]See Samvelian (2012) for an application of this analysis to the CPrs formed with *zadan* 'to hit'. See also Müller (2010) for a partially comparable approach within the HPSG framework.

non-compositional Constructions (or classes) can also be productive. The productivity of Persian CPrs is also related to other parameters such as the coherence of the classes and their size.

4.4 Concluding remarks on complex predicates

Like the *Ezāfe* construction and DOM, Persian CPrs have still a lot to reveal on the many faces of predicate formation in languages of the world and key issues such as the idiomaticity vs. compositionality or storage vs. online processing. Here again, there is a lot to gain from resorting to empirical methods, which can provide a new insight to elucidate some crucial theoretical issues discussed since the late 1990s concerning CPr formation in Persian. The issue of the productivity of Persian CPrs, for instance, cannot be adequately investigated without taking into account data from usage and without resorting to quantitative methods comparable to those used in morphology (Baayen 1992). Likewise, the issue of whether Persian CPrs must be stored in the (mental) lexicon cannot receive a valid answer without psycholinguistic investigation (Baayen 2007). Recent studies such as Shabani-Jadidi (2015) and Sadat Safavi et al. (2016) have opened the way for other studies to come.

5 Conclusion

The main purpose of this article was to offer an overview of the issues raised by three specific features of Persian syntax, namely the *Ezāfe* construction, differential object marking, and complex predicate formation, and the way

various studies have tried to account for these issues. Because of space limitations and the impressive number of studies, it was impossible to get into the details and subtleties of all studies presented through the article. Hopefully, nothing has been "lost in translation" and the quoted authors' positions have been rendered faithfully.

Noting the enduring interest for these three phenomena in the literature since the early 1980s, one may (wrongly) assume that they have disclosed all their secrets and that almost everything worth saying has already been said. Along with the presentation of the huge amount of work already done, another aim of this article was to show that each of three phenomena at stake still offers a challenging and promising area for empirical and theoretical investigation, as illustrated by a number of recent studies adopting new methodological approaches.

References

Aissen, J. (2003), 'Differential Object Marking: Iconicity vs. Economy', *Natural Language and Linguistic Theory* **21**(3), 435–83.

Anderson, S. R. (1971), 'On the Role of Deep Structure in Semantic Interpretation', *Foundations of Language* **7**, 387–96.

Aronoff, M. (1993), *Morphology by Itself*, MIT Press, Cambridge, MA.

Baayen, R. H. (1992), Quantitative aspects of morphological productivity, *in* G. E. Booij & J. van Marle, eds, 'Yearbook of Morphology 1991', Kluwer Academic Publishers, Dordrecht, pp. 109–49.

Baayen, R. H. (2007), Storage and computation in the mental lexicon, *in* G. Jarema & G. Libben, eds, 'The mental lexicon: Core perspectives', Elsevier, Amsterdam, pp. 81–104.

Barjasteh, D. (1983), Morphology, syntax and Semantics of Persian Compound Verbs: A Lexical Approach, PhD thesis, University of Illinois.

Bashiri, I. (1981), *Persian Syntax*, Burgess Publishing Company, Minneapolis.

Bonami, O. & Samvelian, P. (2010), 'Persian complex predicates: Lexeme formation by itself.', Paper presented at Septièmes Décembrettes Morphology Conference, Toulouse.

Bonami, O. & Samvelian, P. (2015), 'The diversity of inflectional periphrasis in persian', *Journal of Linguistics* **51**(1), 327–382.

Borer, H. (1994), The Projection of Arguments, *in* E. Benedicto & J. T. Runner, eds, 'Functional Projections. University of Massachusetts Occasional Papers in Linguistics 17', University of Massachusetts and GLSA Publications, Amherst.

Bossong, G. (1985), *Empirische Universalienforschung. Differentielle Objektmarkierung in der neuiranischen Sprachen*, Narr, Tübingen.

Bossong, G. (1991), Differential object marking in Romance and beyond, *in* D. Wanner & D. A. Kibbee, eds, 'New Analyses in Romance Linguistics', John Benjamins, Amsterdam, pp. 143–167.

Bresnan, J. (2006), Is syntactic knowledge probabilistic? Experiments with the English dative alternation, *in* S. Featherston & W. Sternefeld, eds, 'Roots: Linguistics in search of its evidential base', Mouton de Gruyter, Berlin, pp. 75–96.

Bresnan, J., Cueni, A., Nikitina, T. & Baayen, R. H. (2007), Predicting the dative alternation, *in* G. Bouma, I. Kraemer & J. Zwarts, eds, 'Cognitive Foundations of Interpretation', Royal Netherlands Academy of Science, Amsterdam, pp. 69–94.

Bresnan, J. & Hay, J. (2008), 'Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English', *Lingua* **118**(2), 245–59.

Browne, W. (1970), 'More on definiteness markers: Interrogatives in Persian', *Linguistic Inquiry* **1**(3), 359–63.

Browning, M. & Karimi, E. (1994), Scrambling to object position in persian, *in* H. van Riemsdijk & N. Corver, eds, 'Studies in Scrambling. Movement and Non-Movement Approaches to Free Word-Order Phenomena', Mouton de Gruyter, Berlin and New York, pp. 61–100.

Bubenik, V. (2009), The rise and development of the possessive construction in Middle Iranian with parallels in Albanian, *in* V. Bubenik, J. Hewson & S. Rose, eds, 'Grammatical Change in Indo-European Languages', Current issues in linguistic theory, John Benjamins, Amsterdam and Philadelphia, pp. 89–102.

Butt, M. (1995), *The Structure of Complex Predicates*, CSLI Publications, Stanford.

Chomsky, N. (1970), Remarks on nominalization, *in* R. A. Jacobs & P. S. Rosenbaum, eds, 'Readings in English Transformational Grammar', Ginn and Co, Waltham, MA, pp. 184–221.

Chomsky, N. (1981), *Lectures on Government and Binding*, Foris, Dordrecht.

Chomsky, N. (1995), *The Minimalist Program*, MIT Press, Cambridge, MA.

Cinque, G. (2005), 'Deriving Greenberg's universal 20 and its exceptions', *Linguistic Inquiry* **36**, 315–332.

Cinque, G. (2010), *The Syntax of Adjectives: A Comparative Study*, MIT Press, Cambridge, MA.

Comrie, B. (1979), 'Definite and animate direct objects: A natural class', *Linguistica Silesiana* **3**, 13–21.

Comrie, B. (1989), *Language Universals and Linguistic Typology*, University of Chicago Press, Chicago.

Dabir-Moghaddam, M. (1990), 'Pirāmun-e rā dar zabān-e fārsi [On *rā* in Persian]', *Majalleye Zabānšenāsi* **13**(1), 2–60.

Dabir-Moghaddam, M. (1992), On the (in)dependence of syntax and pragmatics: Evidence from the postposition *rā* in persian, *in* D. Stein, ed., 'Cooperating with Written Texts: The Pragmatics and Comprehension of Written Texts', Mouton de Gruyter, Berlin, pp. 549–73.

Dabir-Moghaddam, M. (1995), 'Fe'le morakkab dar zabāne fārsi [Compound verbs in Persian]', *Majalleye Zabānshenāsi* **12**, 2–46.

Dabir-Moghaddam, M. (1997), 'Compound verbs in persian', *Studies in the Linguistic Sciences* **27**(2), 25–59.

Dabir-Moghaddam, M. (2009), Zabāne fārsi va nazariyehāye zabāni: dar jostojuye čarčubi barāye tadvin dasture zabān-e jāme'e za bāne fārsi, *in* 'Pažuheshāye zabānšenāxtiye fārsi', Markaze našre dānešgāhi, Tehran, pp. 230–67.

Dalrymple, M. & Nikolaeva, I. (2011), *Objects and Information Structure*, Cambridge University Press, Cambridge.

Darmesteter, J. (1883), *Études sur la grammaire historique de la langue persane*, F. Vieweg, Paris.

de Swart, P. (2007), Cross-linguistic Variation in Object Marking, PhD thesis, Radboud University, Nijmegen.

Den Dikken, M. (2006), *Relators and Linkers: A Study of Predication, Predicate Inversion, and Copulas*, The MIT Press, Cambridge, MA.

Di Sciullo, A.-M. & Williams, E. (1987), *On the Definition of Word*, MIT Press, Cambridge, MA.

Diesing, M. (1992), *Indefinites*, The MIT Press, Cambridge, MA.

Enç, M. (1991), 'The semantics of specificity', *Linguistic Inquiry* **22**, 1–26.

Estaji, A. (2009), Morphosyntactic changes in Persian and their effects on syntax, *in* V. Bubenik, J. Hewson & S. Rose, eds, 'Grammatical Change in Indo-European Languages', John Benjamins, Amsterdam, pp. 195–206.

Faghiri, P. (2016), Une étude empirique de l'ordre des mots dans le domaine préverbal en persan, PhD thesis (forthcoming), Université Sorbonne Nouvelle.

Faghiri, P. & Samvelian, P. (2014), Constituent ordering in Persian and the weight factor, *in* C. Pinon, ed., 'Empirical Issues in Syntax and Semantics 10', pp. 215–232.

Faghiri, P. & Samvelian, P. (2015*a*), How much is determined by syntax? An empirical approach to the position of the dialect object in Persian, *in* 'Proceedings of the 9th Iranian Conference on Linguistics', Allameh Tabatabai University, Tehran, pp. 1419–35.

Faghiri, P. & Samvelian, P. (2015*b*), Jāygāh-e maf'ul-e mostaqim dar fārsi: barresi-e dādebonyād [The position of the direct object in Persian: A corpus-based approach], *in* M. Rezvanian, ed., 'Proceedings of the First National Conference on Semantics and Syntax of Iranian Languages', Tehran University Press, Tehran, pp. 61–74.

Faghiri, P., Samvelian, P. & Hemforth, B. (2014), Accessibility and word order: The case of ditransitive constructions in Persian, *in* S. Müller, ed., 'Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar', CSLI Publications, Stanford, CA, pp. 217–237.

Family, N. (2006), Explorations of Semantic Space: The Case of Light Verb Constructions in Persian, PhD thesis, EHESS.

Family, N. (2009), Mapping semantic spaces: A constructionist account of the "light verb" *eat* in Persian, *in* M. Vanhove, ed., 'From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations', John Benjamins, Amsterdam and Phildelphia.

Family, N. (2014), *Semantic Spaces of Persian Light Verbs: A Constructionist Account*, Brill's Studies in South and Southwest Asian Languages, Brill, Leiden.

Farkas, D. (1995), Specificity and scope, *in* L. Nash & G. Tsoulas, eds, 'Actes du Premier Colloque Langues et Grammaire', Presses Universitaires de Vincennes, Paris, pp. 119–37.

Farkas, D. (2002), 'Specificity distinction', *Journal of Semantics* **19**, 213–243.

Fillmore, C. J., Kay, P. & O'Connor, M. C. (1988), 'Regularity and idiomaticity in grammatical constructions: The case of *let alone*', *Language* **64**(3), 501–38.

Fodor, J. & Sag, I. A. (1982), 'Referential and quantificational indefinites', *Linguistics and Philosophy* **5**, 355–98.

Folli, R., Harley, H. & Karimi, S. (2005), 'Determinants of event type in Persian complex predicates', *Lingua* **115**, 1365–1401.

Gaeta, L. & Ricca, D. (2009), 'Composita solvantur: Compounds as lexical units or morphological objects?', *Rivista di Linguistica* **21**(1), 35–70.

Ganjavi, S. (2007), Direct objects in Persian, PhD thesis, University of Southern California.

Ganjavi, S. (2011), On direct objects in Persian: The case of the non-rā marked DOs, *in* A. Korn, G. Haig, S. Karimi & P. Samvelian, eds, 'Topics in Iranian Linguistics', Dr. Ludwig Reichert Verlag, Wiesbaden, pp. 167–81.

Ghaniabadi, S. (2010), The Empty Noun Construction in Persian, PhD thesis, University of Manitoba.

Gharib, A., Bahār, M., Foruzānfar, B., Homāyi, J., & Yāssemi, R. (1994), *Dastur-e Zabān-e fārsi. Panj Ostād [Grammar of Persian]*, Entešārāt-e Nāhid, Tehran.

Ghomeshi, G. & Massam, D. (1994), 'Lexical/syntactic relations without projection', *Linguistic Analysis* **24**, 175–217.

Ghomeshi, J. (1996), Projection and Inflection: A Study of Persian Phrase Structure, PhD thesis, University of Toronto.

Ghomeshi, J. (1997*a*), 'Non-projecting nouns and the Ezafe Construction in Persian', *Natural Language & Linguistic Theory* **15**(4), 729–88.

Ghomeshi, J. (1997*b*), 'Topics in Persian VPs', *Lingua* **102**, 133–67.

Givi Ahmadi, H. & Hassan, A. (1995), *Dastur-e zabāne fārsi [A Grammar of Persian]*, Mo'assese-ye Farhangi Fātemi, Tehran.

Givón, T. (1984), Direct object and dative shifting: Semantic and pragmatic case, *in* F. Plank, ed., 'Objects: Towards a Theory of Grammatical Relations', Academic Press, London, pp. 151–82.

Goldberg, A. E. (1995), *A Construction Grammar Approach to Argument Structure*, University of Chicago Press, Chicago.

Goldberg, A. E. (1996), Words by default: Optimizing constraints and the Persian complex predicate, *in* 'Proceedings of the Berkeley Linguistic Society 22', Berkeley, pp. 132–146.

Goldberg, A. E. (2003), Words by default: The Persian complex predicate construction, *in* E. J. Francis & L. A. Michaelis, eds, 'Mismatch: Form-Function Incongruity and the Architecture of Grammar', CSLI Publications, Stanford, pp. 117–46.

Grimshaw, J. & Mester, A. (1988), 'Light verbs and theta-marking', *Linguistic Inquiry* **19**, 205–32.

Gundel, J. K. (1988), *The Role of Topic and Comment in Linguistic Theory*, Garland, New York.

Haider, H. & Zwanziger, R. (1984), Relatively attributive: The ezāfe construction from Old Iranian to Modern Persian, *in* J. Fisiak, ed., 'Historical Syntax', Mouton, Berlin, pp. 137–72.

Haig, G. L. J. (2011), Linker, relativizer, nominalizer, tense-particle: On the ezāfe in West Iranian, *in* F. H. Yap, K. Grunow-Harsta & J. Wrona, eds,

'Nominalization in Asian Languages. Diachronic and Typological Perspectives', John Benjamins, Amsterdam, pp. 363–90.

Hale, K. L. & Keyser, S. J. (1992), The syntactic character of thematic structure, *in* I. M. Roca, ed., 'Thematic Structure: Its Role in Grammar', Foris, Berlin, pp. 107–43.

Hale, K. L. & Keyser, S. J. (1993), On argument structure and the lexical expression of syntactic relations, *in* K. Hale & S. J. Keyser, eds, 'The View from Building 20', MIT Press, Cambridge, MA, pp. 53–109.

Hale, K. L. & Keyser, S. J. (1997), On the complex nature of simple predicators, *in* A. Alsina, J. Bresnan & P. Sells, eds, 'Complex Predicates', CSLI Publications, Stanford, pp. 29–65.

Hale, K. L. & Keyser, S. J. (2002), *Prolegomenon to a Theory of Argument Structure*, MIT Press, Cambridge, MA.

Hashemipour, M. M. (1989), Pronominalization and control in modern Persian, Doctoral dissertation, University of California, San Diego.

Hincha, G. (1961), 'Beiträge zu einer Morphemlehre des Neupersischen', *Der Islam* **37**, 136–201.

Hintikka, J. (1986), 'The semantics of 'a certain'', *Linguistic Inquiry* **17**, 331–36.

Holmberg, A. & Odden, D. (2008), The noun phrase in Hawrami, *in* S. Karimi, V. Samiian & D. Stilo, eds, 'Aspects of Iranian Linguistics', Cambridge Scholars Publishing, Newcastle, pp. 129–52.

Hopper, P. & Thompson, S. (1980), 'Transitivity in grammar and discourse', *Language* **56**, 251–299.

Jackendoff, R. S. (1977), *X-bar-Syntax: A Study of Phrase Structure*, MIT Press, Cambridge, MA.

Jügel, T. (2015), *Die Entwicklung der Ergativkonstruktion im Alt- und Mitteliranischen – Eine korpusbasierte Untersuchung zu Kasus, Kongruenz und Satzbau*, Harrassowitz, Wiesbaden.

Kahnemuyipour, A. (2014), 'Revisiting the Persian ezafe construction: A roll-up movement analysis', *Lingua* **150**, 1–24.

Karimi-Doostan, G. (1997), Light Verb Constructions in Persian, PhD thesis, University of Essex.

Karimi-Doostan, G. (2005), 'Light verbs and structural case', *Lingua* **115**, 1737–56.

Karimi, S. (1989), Aspects of Persian Syntax, Specificity and the Theory of Grammar, PhD dissertation, University of Washington.

Karimi, S. (1990), 'Obliqueness, specificity, and discourse functions: *Rā* in Persian', *Linguistic Analysis* **20**, 139–91.

Karimi, S. (1996), 'Case and specificity: Persian *rā* revisited', *Linguistic Analysis* **26**, 174–94.

Karimi, S. (1997), 'Persian complex verbs: Idiomatic or compositional', *Lexicology* **3**, 273–318.

Karimi, S. (1999), 'A note on parasitic gaps and specificity', *Linguistic Inquiry* **30**, 704–13.

Karimi, S. (2003), On object positions, specificity and scrambling in Persian, *in* S. Karimi, ed., 'Word Order and Scrambling', Blackwell, Oxford, pp. 91–124.

Karimi, S. & Brame, M. (2012), 'A generalization concerning the Ezafe construction in Persian', *Linguistic Analysis* **38**, 111–43.

Karimi, S. & Smith, R. (2015), The mystery of persian –râ. ms., university of arizona. Ms., University of Arizona.

Karimi, Y. (2007), 'Kurdish ezafe construction: Implications for DP structure', *Lingua* **117**, 2159–77.

Kay, P. & Fillmore, C. J. (1999), 'Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction', *Language* **75**, 1–33.

Kent, R. (1953), *Old Persian. Grammar, Texts, Lexicon*, 2nd revised edition edn, American Oriental Society, New Haven.

Kent, R. G. (1944), 'The Old Persian relative and article', *Language* **20**, 1–10.

Khānlari, P. (1972), *Dasture zabāne fārsi [Grammar of Persian]*, Bonyād-e farhang, Tehran.

Khānlari, P. (1986), *Tarix-e zabn̄-e farsi (A History of the Persian Language)*, Editions Nashr-e Now.

Klavans, J. L. (1985), 'The independence of syntax and phonology in cliticization', *Language* **61**(1), 95–120.

Korn, A. (2013), 'Looking for the middle way: Voice and transitivity in complex predicates in iranian', *Lingua* **135**, 30–55.

Kratzer, A. (1996), Severing the external argument from its verb, *in* J. Rooryck & L. Zaring, eds, 'Phrase Structure and the Lexicon', Kluwer Academic Publishers, Dordrecht, pp. 109–37.

Lambrecht, K. (1994), *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*, Cambridge University Press, Cambridge.

Lapointe, S. G. (1990), Edge features in GPSG, *in* M. Ziolkowski, M. Noske & K. Deaton, eds, 'Papers from the 26th Regional Meeting of the Chicago Linguistic Society', Chicago Linguistic Society, Chicago, pp. 221–35.

Lapointe, S. G. (1992), Life on the edge: argument in favor of an Autolexical account of edge inflections, *in* C. P. Canakis, G. P. Chan & J. M. Denton, eds, 'Papers from the 28th Regional Meeting of the Chicago Linguistic Society', Chicago Linguistic Society, Chicago, pp. 318–33.

Larson, R. (1988), 'On the double object construction', *Linguistic Inquiry* **19**, 335–91.

Larson, R. & Yamakido, H. (2008), Ezafe and the deep position of nominal modifiers, *in* L. McNally & C. Kennedy, eds, 'Adjectives and Adverbs: Syntax, Semantics and Discourse', Oxford University Press, Oxford, pp. 43–70.

Lazard, G. (1957), *Grammaire du persan contemporain*, Klincksieck, Paris.

Lazard, G. (1966), 'L'enclitique nominal -*i* en persan : Un ou deux morphèmes?', *Bulletin de la Société de Linguistique de Paris* **61**, 249–64.

Lazard, G. (1982), 'Le morphème *rā* en persan et les relations actancielles', *Bulletin de la Société de Linguistique de Paris* **77**, 177–207.

Lazard, G. (1984), Actance variations and categories of the object, *in* F. Plank, ed., 'Objects: Towards a Theory of Grammatical Relations', Academic Press, London, pp. 269–92.

Lazard, G. (1994), 'Le *rā* persan et le *ba* chinois', *Cahiers de linguistique Asie orientale* **23**(1), 169–76.

Lazard, G. (2013), 'Le verbe persan *zadan* 'frapper', ses constructions et ses emplois', *Studia Iranica* **42**(1), 7–21.

Lecoq, P. (2002), *Recherches sur les Dialectes Kermaniens*, Vol. 39 of *Acta Iranica*, Peeters, Leuven.

Levin, B. (1993), *English Verb Classes and Alternations*, The University of Chicago Press, Chicago.

MacKenzie, D. N. (1961), *Kurdish Dialect Studies, Vol. I.*, Oxford University Press, London.

Mackenzie, D. N. (1966), *The Dialect of Awroman. Grammatical Sketch, Texts, and Vocabulary*, Munksgaard, Copenhagen.

Mahootian, S. (1997), *Persian*, Routledge, London and New York.

Malchukov, A. L. (2008), 'Animacy and asymmetries in differential case marking', *Lingua* **118**(2), 203–21.

Marantz, A. (1997), 'No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon', *U. Penn Working Papers in Linguistics* **4.2**, 201–225.

Masini, F. (2009), 'Phrasal lexemes, compounds and phrases: a constructionist perspective', *Word Structure* **2**(2), 254–71.

Megerdoomian, K. (2001), 'Event structure and complex predicates in persian', *Canadian Journal of Linguistics* **46**, 97–125.

Megerdoomian, K. (2002), Beyond Words and Phrases: A Unified Theory of Predicate Composition, PhD dissertation, University of Southern California.

Megerdoomian, K. (2012), 'The status of the nominal in Persian complex predicates', *Natural Language and Linguistic Theory* **30**(1), 179–216.

Meillet, A. (1931), *Grammaire du vieux-perse*, Honoré Champion, Paris.

Meunier, A. & Samvelian, P. (1997), 'La postposition *rā* en persan : son rôle dans la détermination et sa fonction discursive', *Cahiers de grammaire* **22**, 187–231.

Miller, P. H. (1992), *Clitics and Constituents in Phrase Structure Grammar*, Garland, New York.

Modarresi, F. (2014), Bare Nouns in Persian: Interpretation, Grammar, and Prosody, PhD dissertation, University of Ottawa.

Mohammad, J. & Karimi, S. (1992), Light verbs are taking over: Complex verbs in Persian, *in* 'Proceedings of the Western Conference on Linguistics (WECOL) 5', pp. 195–212.

Müller, S. (2010), 'Persian complex predicates and the limits of inheritance-based analyses', *Journal of Linguistics* **46**(2), 601–55.

Næss, Å. (2007), *Prototypical Transitivity*, John. Benjamins, Amsterdam.

Næss, r. (2004), 'What markedness marks: The markedness problem with direct objects', *Lingua* **114**, 1186–1212.

Nichols, J. (1986), 'Head-marking and dependent-marking languages', *Language* **62**, 56–119.

Nikolaeva, I. (2001), 'Secondary topic as a relation in information structure', *Linguistics* **39**, 1–49.

Nunberg, G., Sag, I. A. & Wasow, T. (1994), 'Idioms', *Language* **70**, 491–538.

Palmer, A. (1971), The EZAFE Construction in Modern Standard Persian, PhD dissertation, University of Michigan.

Pantcheva, M. (2010), 'First phase syntax of persian complex predicates: Argument structure and telicity', *Journal of South Asian Linguistics* **2**(1), 53–72.

Paul, L. (1998), *Zazaki. Grammatik und Versuch einer Dialektologie.*, Dr. Ludwig Reichert Verlag, Wiesbaden.

Paul, L. (2003), Early Judaeo-Persian in a historical perspective: The case of the prepositions *be*, *u*, *pa(d)*, and the suffix *rā*, *in* L. Paul, ed., 'Persian Origins - Early Judaeo-Persian and the Emergence of New Persian', Harrassowitz Verlag, Wiesbaden, pp. 177–94.

Perry, J. (2005), *A Tajik Persian Reference Grammar.*, Brill, Leiden.

Pesetsky, D. M. (1985), 'Morphology and logical form', *Linguistic Inquiry* **16**, 193–246.

Peterson, D. (1974), Noun Phrase Specificity, PhD thesis, University of Michigan.

Phillott, D. (1919), *A Higher Persian Grammar*, Calcutta.

Polinsky, M. (1995), 'Double objects in causatives: Towards a study of coding conflict', *Studies in Language* **19**, 129–221.

Pollard, C. J. & Sag, I. A. (1994), *Head-Driven Phrase Structure Grammar*, University of Chicago Press, Chicago.

Rasekhmahand, M. (2004), 'Jāygāhe maf'ule mostaqim dar fārsi [The position of the direct object in Persian]', *Nāme farhangestān* **6**, 56–66.

Reinhart, T. (1982), 'Pragmatics and linguistics: An analysis of sentence topics', *Philosophica* **27**, 53–94.

Riemsdijk, H. & Williams, E. (1981), 'NP structure', *The Linguistic Review* **1**, 171–217.

Ritter, E. & Rosen, S. (1996), 'Strong and weak predicates: Reducing the lexical burden', *Linguistic Analysis* **26**, 1–34.

Ritter, E. & Rosen, S. T. (2001), 'The interpretive value of object splits', *Language Sciences* **23**, 425–51.

Roberts, J. R., Barjasteh Delforooz, B. & Jahani, C. (2009), *A Study of Persian Discourse Structure*, Acta Universitatis Upsaliensis, Uppsala.

Sadat Safavi, M., Husain, S. & Vasishth, S. (2016), 'Dependency resolution difficulty increases with distance in persian separable complex predicates: Evidence for expectation and memory-based accounts', *Frontiers in Psychology* **7**.

Sadeghi, A. A. (1970), '*Rā* dar zabān-e fārsi-e emrooz [*Rā* in the Persian of today]', *Nashriye-ye dāneshkade-ye adabiyāt va olum-e ensāni-e Tabriz* **93**, 9–22.

Sadeghi, A. A. (1993), On denominative verbs in Persian, *in* 'Farsi Language and the Language of Science', University Press, Tehran, pp. 236–46.

Samiian, V. (1983), Origins of Phrasal Categories in Persian: An X-bar Analysis, PhD dissertation, UCLA.

Samiian, V. (1994), The Ezafe construction: Some implications for the theory of X-bar syntax, *in* M. Marashi, ed., 'Persian Studies in North America', Iranbooks, Bethesda, pp. 17–41.

Samsam Bakhtiari, F. (2000), La composition dans la relation verbe-objet.

Problématique générale et application au persan, Thèse de Doctorat, Université Lumière Lyon 2.

Samvelian, P. (2001), 'Le statut syntaxique des objets nus en persan', *Bulletin de la Société de Linguistique de Paris* **96**(1), 349–88.

Samvelian, P. (2002), Une analyse non configurationnelle du morphème *rā* en persan. Unpublished manuscript.

Samvelian, P. (2006*a*), 'Le sujet, l'objet et l'inaccusativité dans les prédicats complexes nom-verbe en persan', *Cahiers de linguistique de l'INALCO* **6**, 155–90.

Samvelian, P. (2006*b*), 'L'enclitique -*i* introducteur de relative en persan: déterminant, allomorphe de l'*ezâfé* ou autre chose encore?', *Studia Iranica* **35**(1), 7–34.

Samvelian, P. (2007), 'A (phrasal) affix analysis of the Persian Ezafe', *Journal of Linguistics* **43**, 605–45.

Samvelian, P. (2008), The Ezafe as a head-marking inflectional affix: Evidence from Persian and Kurmanji Kurdish, *in* V. S. Simin Karimi & D. Stillo, eds, 'Aspects of Iranian Linguistics: Papers in Honor of Mohammad Reza Bateni', Cambridge Scholars Ltd, Cambridge, pp. 339–61.

Samvelian, P. (2012), *Grammaire des prédicats complexes. Les constructions nom-verbe*, Lavoisier, Paris.

Samvelian, P. & Faghiri, P. (2013*a*), Introducing PersPred, a syntactic and semantic database for Persian complex predicates, *in* 'Proceedings of the 9th

Workshop on Multiword Expressions (MWE 2013), NAACL-HLT, June 2013', Association for Computational Linguistics, Atlanta, pp. 11–20.

Samvelian, P. & Faghiri, P. (2013*b*), Re-thinking compositionality in Persian complex predicates, *in* 'Proceedings of the 39th Berkeley Linguistics Society', Linguistic Society of America, Berkeley, p. Forthcoming.

Samvelian, P. & Faghiri, P. (2014), 'Persian complex predicates: How compositional are they?', *Semantics-Syntax Interface* **1**, 43–74.

Schroeder, C. (1999), Attribution in Kurmanji (Nordkurdisch), *in* K. H. Wagner & W. Wildgen, eds, 'Studien zur Phonologie, Grammatik, Sprachphilosophie and Semiotik', Institüt für Allgemeine und Angewandte Sprachwissenschaft, Bremen, pp. 43–63.

Sedighi, A. (2009), 'Do psychological constructions in Persian involve complex predicates?', *Rice Working Papers in Linguistics* **1**, 65–78.

Shabani-Jadidi, P. (2015), *Processing Compound Verbs in Persian: A Psycholinguistic Approach to Complex Predicates*, Leiden University Press, Leiden.

Shokouhi, H. & Kipka, P. (2003), 'A discourse study of Persian *rā*', *Lingua* **113**(10), 953–66.

Spencer, A. (1988), 'Bracketing paradoxes and the English lexicon', *Language* **64**, 663–82.

Sproat, R. (1984), 'On bracketing paradoxes', *MIT Working Papers in Linguistics* **7**, 110–30.

Tabaian, H. (1979), 'Persian compound verbs', *Lingua* **47**(2-3), 189–208.

Telegdi, Z. (1951), 'Nature et fonction des périphrases verbales dites "verbes composés" en persan', *Acta Orientalia Hungarica* **1**, 315–38.

Tseng, J. (2003), Phrasal affixes and French mophosyntax, *in* G. Jäger, P. Monachesi, G. Penn & S. Wintner, eds, 'Proceedings of Formal Grammar 2003 (FGVienna)', CSLI Publications, Stanford, CA, pp. 177–88.

Vahedi-Langrudi, M.-M. (1996), The Syntax, Semantics and Argument Structure of Complex Predicates in Modern Farsi, PhD thesis, University of Ottawa.

van Geenhoven, V. (1998), *Semantic Incorporation and Indefinite Descriptions*, CSLI Publications, Stanford.

Wasow, T. (2002), *Postverbal Behavior*, CSLI Publications, Stanford, CA.

Williams, E. (1981), 'On the notions "lexically related" and "head of a word"', *Linguistic Inquiry* **1**, 245–74.

Windfuhr, G. L. (1979), *Persian Grammar: History and State of its Study*, Mouton, The Hague.

Windfuhr, G. L. (1989), New West Iranian, *in* R. Schmitt, ed., 'Compendium Linguarum Iranicarum', Ludwig Reichert Verlag, Wiesbaden, pp. 251–62.

Windfuhr, G. L., ed. (2009), *The Iranian Languages*, Routledge, London and New York.

Zwicky, A. (1987), 'Suppressing the z's', *Journal of Linguistics* **23**, 133–48.

Zwicky, A. & Pullum, G. K. (1983), 'Cliticization vs. inflection: English *n't*', *Language* **59**, 502–13.