

# Persian Complex Predicates: How Compositional Are They?

*Pollet Samvelian*<sup>a\*</sup> and *Pegah Faghiri*<sup>b</sup>

<sup>a,b</sup>Université Sorbonne Nouvelle & CNRS, Paris, France

**Abstract:** In this paper, we address the issue of compositionality of Persian complex predicates. We argue against the “radical” or “fully” compositional views prevailing in the studies within the generative framework. We further show that any *a priori* compositional account of Persian complex predicates is doomed to failure, since the respective contribution of the verb and the preverbal element cannot be determined *a priori* and irrespective of the other member. We claim, nevertheless, that a compositional account of these combinations is possible, provided compositionality is defined *a posteriori* as in idiomatically combining expressions. We then present a Construction-based account that allows to account for the productivity of Persian complex predicates in a way which does justice to their compositionality without overlooking their idiomacity.

**Keywords:** Persian complex predicates, compositionality, idiomacity, Construction Grammar

## 1. Introduction

Although Persian complex predicates (CPr) have been a focus of interest for almost thirty years now, most of theoretical studies have unfortunately explored a limited set of data. As a consequence, their empirical underpinnings are not reliable and their

---

\* Email: pollet.samvelian@univ-paris3.fr

A Part of the work reported here was presented at the 39th *Annual Meeting of the Berkeley Linguistics Society* (Berkeley, 2013). We thank the audience for their questions and suggestions. This work is related to the work package LR4.1 of the Labex EFL, a public grant funded by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (reference: ANR-10-LABX-0083).

generalizations turn out to be inaccurate when a larger set of data is taken into account. The problem is particularly sharp when it comes to the issue of compositionality, one of the most discussed questions in recent studies. Since Karimi-Doostan's (1997) seminal work on light verb constructions in Persian, a "radical" or "fully" compositional view of Persian CPrs has progressively become prevailing in the studies within the generative framework. Despite substantial variations, these studies all rely on the assumption that not only the verb and the non-verbal element of a CPr have each a precise contribution to its meaning but also that their contribution is defined *a priori* and remains consistent through all their combinations to form a CPr. Divergences only rise with respect to the "division of labor" between the components and between the lexicon or on the syntax.

Although these studies do not deny the idiomaticity of Persian CPrs, they nevertheless claim that at least the argument structure and the eventive/aspectual properties of a CPr can be deduced on the basis of its component parts.

In this paper, after discussing one of the recent and most influential "constructionist" analyses of Persian CPrs, namely Folli, Harley and Karimi (2005), we show that none of the generalizations put forward by the authors hold a close scrutiny and that contrary to their claim, the respective contribution of the verb and the preverbal element of a CPr cannot be determined *a priori* and irrespective of the other member. We further claim that any *a priori* compositional account faces similar problems.

We argue, nevertheless, that a compositional account of the Persian CPrs is possible, provided compositionality is defined *a posteriori* as in idiomatically combining expressions in the sense of Nunberg, Sag and Wasow (1994). We then present a Construction-based compositional account of Persian CPrs developed by Samvelian (2012) and Samvelian and Faghiri (in press) and implemented in the *PersPred* database (Samvelian and Faghiri 2013; Samvelian, Faghiri and El Ayari 2014).

## 2. Folli, Harley and Karimi's (2005) Account of Persian CPrs

The compositionality of Persian CPrs may seem strikingly obvious when examples such as those in (1) and (2) are considered:<sup>1</sup>

---

<sup>1</sup> Glosses follow the Leipzig Glossing Rules ([www.eva.mpg.de/lingua/resources/glossing-rules.php](http://www.eva.mpg.de/lingua/resources/glossing-rules.php)). The following non-standard abbreviations are used for clarity: COP = copula, DOM = differential object marking, EZ = Ezafe.

- (1) 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 xor-d*  
 Omid from Sara slap collide-PST.3SG  
 ‘Omid was slapped by Sara.’
- (2) a. *maryam (hamiše) in ettefāq=rā be yād dāšt*  
 Maryam always this event=DOM to memory have.PST.3SG  
 ‘Maryam (always) remembered this event.’ (durative reading)
- b. *maryam (nāgahān) in ettefāq=rā be yād āvar-d*  
 Maryam suddenly this event=DOM to memory bring-PST.3SG  
 ‘Maryam (suddenly) remembered this event.’ (punctual reading)

In these examples:

- (a) The referential meaning of the CPr and the roles assigned to the event participants are determined by the nominal element, since the semantic participants of the CPr, as in (1b), are identical to those of the noun within the NP it projects, as in (1a).
- (b) The verb, on the other hand, determines the argument mapping, since its substitution by another verb entails a change in the mapping of the participants and the grammatical functions, as in (1b) and (1c).
- (c) The verb seems to determine also a part of the aspectual properties of the CPr, since the verb alternation in (2), *dāštan* ‘to have’ vs. *āvardan* ‘to bring’, gives rise to an aspectual contrast.

Building on data comparable to those in (1) and (2), Karimi-Doostan (1997) provides one of the first serious attempts to formalize the respective contribution of the verb and the non-verbal element in CPr formation. Based on Butt’s (1995) work on argument structure, Karimi-Doostan proposes a projectionist account in terms of argument “fusion” or “reformation”. Following Grimshaw and Mester (1988), he assumes that the light verb (LV) does not assign theta-roles and thus 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 rise 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 (1). Non-compositional CPrs are formed when the LV combines with a “thematically opaque” noun, a noun that does not display an argument structure. *yax zadan* ‘to freeze’ (Lit. ‘ice hit’), *qofl 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.

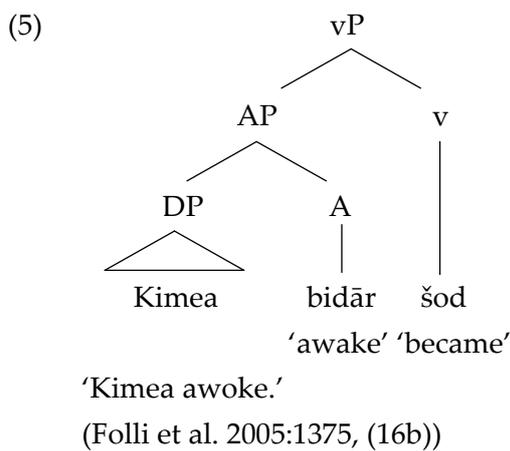
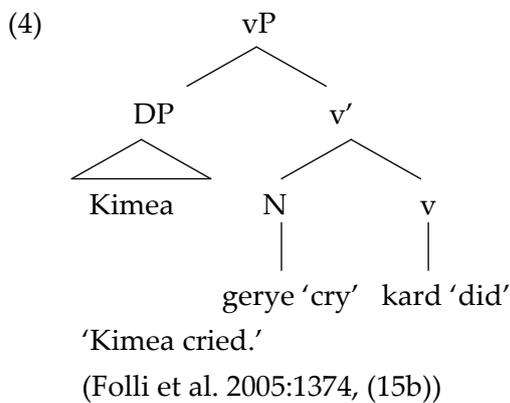
LVs are divided into three categories with respect to their aspectual properties: *Initiatory*, such as *dādan* ‘to give’, *Transition*, such as *xordan* ‘to collide, to strike’ and *Stative*, such as *dāštan* ‘to have’. Some verbs may belong to more than one category, for example *kardan* ‘to do’, which is either *Initiatory* or *Transition*, and thus has 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, that is, either unergative or transitive (or causative) CPrs. They are thus compatible with nouns having at least one external argument, which refers to the initiator of the action denoted by the CPr and is associated to the aspectual role *Initiatory* assigned by the verb. *Transition* verbs form CPrs with a single internal argument, or, in other words, unaccusative predicates. They are thus 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 *šekast* ‘defeat’, which assigns Agent and Patient thematic roles, can either combine with an *Initiatory* verb or a *Transition* verb. In the first case, its external argument (i.e. the Agent) is mapped into the subject function, as in (3a), while in the second case, it is the internal argument that becomes the subject, as in (3b).

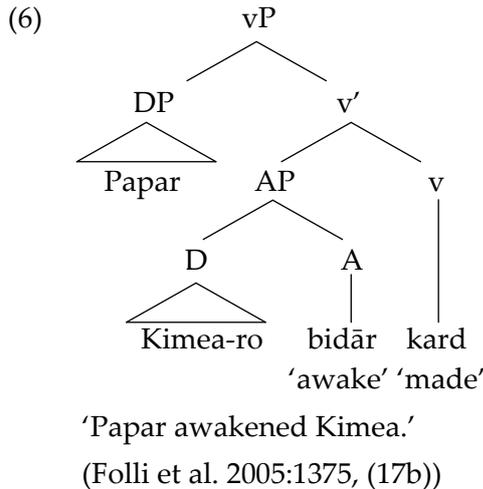
- (3) a. ali sāsān=rā šekast dād  
 Ali Sasan=DOM defeat give.PST.3SG  
 ‘Ali defeated Sasan.’
- b. sāsān az ali šekast xor-d  
 Sasan from Ali defeat collide-PST.3SG  
 ‘Sasan was defeated by Ali.’

As mentioned previously, Karimi-Doostan’s (1997) analysis postulates a categorical distinction between compositional and non-compositional CPrs. Non-compositional CPrs are disregarded, even though, a part of the regularities mentioned above also hold for them.

Megerdooian (2001, 2012) and Folli et al. (2005) are representative examples of constructionist approaches to Persian CPRs. Based on work by Hale and 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.

In Folli et al.'s (2005) approach, the verb in the CPR realizes the *v* head in Hale and Keyser's approach, as illustrated in (4)–(6), corresponding respectively to *gerye kardan* 'to cry' (Lit. 'cry do'), *bidār šodan* 'to awake' (Lit. 'awaken become') and *bidār kardan* 'to awaken' (Lit. 'awaken make'). Persian CPRs are thus the non-incorporated counterpart of verbal constructions suggested by Hale and Keyser.





In this approach, the thematic role of Agent/Cause is assigned by *v* to its external argument (Kratzer 1996; Marantz 1997): *kardan*, in (4) and (6), but not *šodan*, in (5), form an agentive predicate. In other words, in Folli et al.’s approach, 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:

“(...) the Agent-selecting properties of any given light verb are consistent across all Complex Predicates formed with a given LV.” (Folli et al. 2005:1376)

“(...) it is clear that agentivity is a property of the LV in the CPr, and never depends on the nature of the NV element selected.” (Folli et al. 2005:1377)<sup>2</sup>

Note that Megerdooonian (2001) claims the same:

“(...) the choice of the light verb determines whether an external argument is projected.” (Megerdooonian 2001:69)

This claim is supported by the fact that changing the verb in a CPr entails a change in the mapping between the arguments and the grammatical functions, as illustrated in (3).

Following Bashiri (1981), it is claimed that the verb determines also some aspectual properties of the CPr, namely dynamic vs. stative and durative vs. punctual:

“In addition to determining whether the CPr is causative and its external argument is agentive, the light verb distinguishes between eventive and stative CPrs.” (Folli et al. 2005:1378)

“Normally the eventiveness of a complex predicate depends on the light verb involved and not on the non-verbal element.” (Folli et al. 2005:1379)

<sup>2</sup> Folli et al. (2005) notice, however, an exception here: Predicates formed with movement verbs such as *āmadan* ‘to come’ or *gozaštan* ‘to pass’ may be agentive or not, depending on the non-verbal element. This is due to the fact that in most languages, movement verbs may have both, an inchoative/unaccusative or an agentive/unergative reading. The agentivity of the CPr depends on the selected reading of the verb.

“Another property that partially depends on the LV is the duration of the CPr.” (Folli et al. 2005:1379)

This explains the aspectual contrast between (2a) and (2b) above. Like Agent-selection, aspectual/eventive properties of a given LV are consistent through all its combinations to form a CPr. For instance, *dāštan* ‘to have’ is always stative.

The non-verbal element, the authors claim, determines the *aktionsart* properties of the CPr, that is, its telicity, and its referential meaning:

“(…) the telicity of the CPr is dependent on the non verbal element involved, in a very transparent fashion.” (Folli et al. 2005:1374)

More precisely, if the non-verbal element of the CPr is a prepositional phrase, a particle, an adjective or an eventive noun, the CPr is telic, otherwise, that is, if the non-verbal element is a non-eventive noun, the CPr is atelic.

Table 1. Telic and atelic complex predicates (Adapted from Folli et al. 2005:1386, Table 3)

<b>Telic Complex Predicates</b>	
PP + LV:	<i>be donyā āmadan</i> ‘to be born’ (Lit. ‘to world come’), <i>be ātaš kešidan</i> ‘to put on fire (Lit. ‘to fire pull’)
Particle + LV:	<i>kenār āmadan</i> ‘to get along’ (Lit. ‘side come’)
Adjective + LV:	<i>derāz kešidan</i> ‘to lay down’ (Lit. ‘long pull’)
Eventive noun + LV:	<i>šekast xordan</i> ‘to be defeated’ (Lit. ‘defeat collide’), <i>šekast dādan</i> ‘to defeat’ (Lit. ‘defeat give’)
<b>Atelic Complex Predicates</b>	
Non-eventive noun + LV:	<i>dast xordan</i> ‘to get touched’ (Lit. ‘hand collide’) <i>kotak xordan</i> ‘to get beaten’ (Lit. ‘beating hit’) <i>dād zadan</i> ‘to yell’ (Lit. ‘scream hit’) <i>dast andāxtan</i> ‘to mock’ (Lit. ‘hand throw’)

### 3. Why “Radical” Compositional Approaches Are Doomed to Failure

Notwithstanding their differences, “radical” compositional approaches all face the same set of problems, since they build on the wrong assumption that the respective contribution of the components participating in CPr formation is consistent through all their combinations and can be defined *a priori*. As it has been extensively argued in Samvelian (2012), 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 combinations with different verbs. This section provides examples of inconsistencies in the behavior of the verb and the non-verbal element of the CPr (Section 3.1).

Two more issues remain with these approaches. The first one concerns the lexical semantics of the CPr and the second the collocational or idiosyncratic association between the LV and the non-verbal element. As we will see in Section 3.2, in most cases, the precise meaning of a CPr cannot be driven from its component parts, even in the case of semantically transparent and fully compositional CPrs. Furthermore, the association between the LV and the non-verbal element is not always predictable on the basis of the properties of each member (Section 3.3).

### 3.1. The Non-consistent Contribution of the CPr Component Parts

In this section we show that neither the verb nor the non-verbal element has a consistent and invariable contribution through its combinations to form a CPr.

#### 3.1.1. The Non-Consistent Contribution of the Verb

Recall that in Folli et al.'s (2005) analysis, the verb is responsible for Agent-selection and some aspectual properties. At this point, it should be mentioned that the only criterion to which Folli *et al.* (2005) resort is the compatibility of a CPr with the adverb *amdan* 'intentionally'. Agentive CPrs, are compatible with *amdan*, while non-agentive ones are not. Note that this criterion singles out only a part of agentive predicates, namely those whose first argument is human, while, agentive predicates are defined by their capacity to project an external argument (human or not). Thus, appealing to this criterion alone, which wrongfully rules out causative predicates with a non-human first argument, is not satisfactory. Hence, in order to obtain a reliable diagnosis of agentivity, we will use two additional criteria: a) the possibility for the nominal element of the CPr to be *rā*-marked, and b) the availability of a passive counterpart for the CPr.

##### 3.1.1.1. *rā*-Marking of the Nominal Element of the CPr

Persian displays Differential Object Marking (DOM), realized by the enclitic *=rā* and triggered roughly by definiteness (Lazard 1982). Consequently, the possibility for an argument of a given verb to be *rā*-marked amounts to the capacity of the latter to project an external argument (Samvelian 2006–2007).<sup>3</sup>

As for CPrs, several studies (Samiian 1983; Ghomeshi and Massam 1994; Samvelian 2001, 2012; among others) have noted that the nominal element of some CPrs can be modified, quantified, determined and eventually *rā*-marked under some conditions. The relevant fact for our discussion is that *rā*-marking can involve the nominal element of "(semantically) intransitive" (unergative) CPrs, such as *harf zadan*

---

<sup>3</sup> This explains why the analysis of *=rā* as an accusative case-marker is common in the studies with the generative framework (Karimi 1996; Ghomeshi 1997; among others).

'to talk' (Lit. 'talk hit'), *tohmat zadan* 'to accuse' (Lit. 'slander hit') and *aks gereftan* 'to photography' (Lit. 'photo take').<sup>4</sup> Consider now the followings examples:

- (7) a. *maryam harf zad*  
 Maryam speech hit.PST.3SG  
 'Maryam talked.'
- b. *maryam in harf-hā=rā zad*  
 Maryam this speech-PL=DOM hit.PST.3SG  
 'Maryam said these.'
- (8) a. *maryam kār kard*  
 Maryam work do.PST.3SG  
 'Maryam worked.'
- b. *maryam in kār-hā=rā kard*  
 Maryam thihs work-PL=DOM do.PST.3SG  
 'Maryam did these works.'
- (9) a. *maryam az omid aks gereft*  
 Maryam from Omid photo take.PST.3SG  
 'Maryam took a picture/pictures of Omid.'
- b. *maryam aks=e avval=rā az omid gereft*  
 Maryam photo=EZ first=DOM from Omid take.PST.3SG  
 'Maryam took the first picture of Omid.'

Although the CPRs in these examples combine with only one direct argument, namely the one realized as the subject, and as such are semantically "intransitive" (or unergative), the nominal element can nevertheless be *rā*-marked. This is possible because these CPRs project an external argument, or, in other words, are not unaccusative. Building on extensive data, Samvelian (2006–2007) shows that in unaccusative CPRs, the nominal element can never be *rā*-marked.

---

<sup>4</sup> By "semantically intransitive" we mean CPRs that select only one semantic argument, the subject of the CPR. Consequently, the LV combines with one nominal (non-prepositional and non-subject) element, which is the nominal element of the CPR itself, as opposed to "transitive" CPRs, like *rang zadan* 'to paint' and *āb dādan* 'to water', in which the LV combines with two nominal (non-prepositional and non-subject) elements. In the case of the latter, the nominal element of the CPR cannot be *rā*-marked.

- (10) a. āb yax bast  
 water ice close.PST.3SG  
 ‘The water froze.’
- b. \*āb in yax=rā bast  
 water this ice=DOM close.PST.3SG

To sum up, the possibility for the nominal element of a CPr to be *rā*-marked constitutes a sufficient condition for considering the CPr as unergative or causative and thus projecting an external argument.<sup>5</sup>

### 3.1.1.2. Availability of a Passive CPr Counterpart

In Persian, the passive voice is realized by a periphrastic construction, whose properties are parallel to those of copula-predicative complement constructions (Moyne 1974; Dabir-Moghaddam 1982; Bonami and Samvelian, in press). This construction involves the past participle of the lexical verb and the auxiliary *šodan* ‘become’.

- (11) a. maryam in ketāb=rā xar-id  
 Maryam this book=DOM buy-PST.3SG  
 ‘Maryam bought this book.’
- b. in ketāb (tavassot=e maryam) xar-id-e šo-d  
 this book by=EZ Maryam buy-PST-PTCP become-PST.3SG  
 ‘This book was bought by Maryam.’

In the case of CPrs, the valency alternation corresponding to the passive alternation can be realized in two ways, either by changing the verb of the CPr, as in (3) above, or by realizing the periphrastic passive form of the verb in the CPr.<sup>6</sup> Note that these two possibilities are not mutually exclusive.

<sup>5</sup> Note that the reverse is not true, that is, the fact that the nominal element of a CPr cannot be *rā*-marked does not necessarily imply that the CPr is unaccusative, since *rā*-marking depends on a variety of conditions a part of which are semantic or pragmatic.

<sup>6</sup> It must be mentioned that Folli et al. (2005) categorically reject this second option, claiming that verb alternation is the only available option to form the passive variant of a CPr (pp. 1395–1396). Samvelian (2012) gives however several attested counterexamples. Here are some additional examples from Bijankhan Corpora (<http://ece.ut.ac.ir/dbrg/bijankhan/>):

- (i) 22 nahang az marg=e goruhi nejāt dād-e šo-d-and  
 22 whale from death=EZ collective rescue give.PST-PTCP become-PST-3PL  
 ‘22 whales were rescued from a collective death.’

- (12) a. *maryam be omid tohmat zad*  
 Maryam to Omid accusation hit.PST.3SG  
 ‘Maryam accused Omid.’
- b. *be omid tohmat zad-e šo-d*  
 to Omid accusation hit.PST-PTCP become-PST.3SG  
 ‘Omid was accused.’
- c. *be omid tohmat xor-d*  
 to Omid accusation collide-PST.3SG  
 ‘Omid was accused.’

As these examples show, even intransitive (unergative) CPrs can be passivized. This is not surprising, since, as Samvelian (2012) extensively argues, the nominal element of a CPr displays all properties of a syntactic complement, namely a direct object. Unaccusative CPrs, on the contrary, cannot be passivized, (Samvelian 2006–2007).

Now we have three criteria at our disposal in order to decide whether a CPr projects an external argument or not. As for aspectual properties, we keep the criteria used by Folli et al., which are mainly those of Vendler (1957). In what follows, we will provide examples of inconsistency in the contribution of a few LVs among the most productive in CPr formation. We should insist that these inconsistencies are not exceptional, as extensively documented by Samvelian (2012).

• The verb *zadan* ‘to hit’

For Folli et al. (2005), *zadan* is consistently agentive and is used as an example to support the consistency of LVs with respect to Agent-selection. However, although a great number of CPrs formed with *zadan* are indeed agentive, e.g. *jāru zadan* ‘to broom’ (Lit. ‘broom hit’) and *tohmat zadan* ‘to slander’ (Lit. ‘slander hit’), there are numerous examples of unaccusative CPrs formed with *zadan*, as illustrated by the following examples:

- (13) a. *in nān kapak zad-e ast*  
 this bread mold hit.PST-PTCP be.PRS.3SG

---

(ii) *525 hektār be kāšt=e sibzamini extesās dād-e šo-d*  
 525 hectare to cultivation=EZ potato assigning give.PST-PTCP become-PST.3SG  
 ‘525 hectares were assigned to the cultivation of potatoes.’

(iii) *agar dar vaqt-e ezāfi hič gol=i zad-e na-šav-ad (...)*  
 if in time=EZ extra no goal=INDF hit.PST-PTCP NEG-become.PRS-3SG (...)  
 ‘If no goals are scored in the extra time...’

'This bread has gone moldy.'

- b. in dar=e āhani zud zang zad  
 this door=EZ iron-made early rust hit.PST.3SG  
 'This iron-made door quickly rusted.'

- c. āb yax zad  
 water ice hit.PST.3SG  
 'The water froze.'

The subject of the CPrs involved in these examples is clearly not an Agent, but a Patient and the Locus of the process denoted by the CPr. The three criteria mentioned previously uncontroversially single out these CPrs as unaccusative:

- *amdan* 'on purpose' is not compatible with these predicates:

- (14) a. #in nān amdan kapak zad-e ast  
 this bread on-purpose mold hit.PST-PTCP be.PRS.3SG  
 '#This bread has gone moldy on purpose.'
- b. #in dar=e āhani amdan zud zang zad  
 this door=EZ iron-made on-purpose early rust hit.PST.3SG  
 '#This iron-made door quickly rusted on purpose.'
- c. #āb amdan yax zad  
 water on-purpose ice hit.PST.3SG  
 '#The water froze on purpose.'

- The nominal element can *never* be *rā*-marked:

- (15) a. \*nān kapak-hā=ye bozorg=rā diruz zad  
 bread mold-pl=EZ big=DOM yesterday hit.PST.3SG
- b. \*dar=e āhani in zang=rā zad  
 door=EZ iron-made this rust=DOM hit.PST.3SG
- c. \*āb in yax=rā zad  
 water this ice=DOM hit.PST.3SG

- The passive construction with *šodan* 'become' is also excluded: \**kapak zade šodan*, \**zang zade šodan*, \**yax zade šodan*.

Consequently, not all CPrs formed with *zadan* are agentive.<sup>7</sup> The same inconsistency is observed with respect to aspectual properties. Remember that in Folli et al.'s analysis, the verb is also responsible for durative vs. punctual and stative vs. dynamic distinctions. Once again, several examples go against this generalization:

- Dynamic CPrs: *jāru zadan* 'to broom', *sili zadan* 'to slap', *qadam zadan* 'to walk'.
- Stative CPrs: *barq zadan* 'to be shiny', *laq zadan* 'to be unsteady'.
- Durative CPrs: *jāru zadan* 'to broom', *qadam zadan* 'to walk'.
- Punctual CPrs: *sili zadan* 'to slap', *hads zadan* 'to guess'.

• The verb *kardan* 'to do'

Folli and al. (2005) consider *kardan* as the lexical counterpart of an abstract *v*, namely DO or CAUSE, which consequently projects an external argument, e.g. *kār kardan* 'to work' (Lit. 'work do'), *parvāz kardan* 'to fly' (Lit. 'fly do') and *rang kardan* 'to paint' (Lit. 'paint do'). However, as for *zadan*, there are non-agentive CPrs formed with *kardan*, as illustrated by *gir kardan* 'to get caught' (Lit. 'blocking do'), *ādat kardan* 'to get accustomed' (Lit. 'habit do'), *rošd kardan* 'to grow' (Lit. 'growing do'), *boruz kardan* 'to reveal oneself' (Lit. 'revelation do'):

- (16) a. mu-hā=ye    maryam    lā=ye            zip    gir            kard  
           hair-pl=EZ    Maryam    between=EZ    zip    blocking    do.PST.3SG  
           'Maryam's hair was caught in the zip.'
- b. maryam    kamkam            be    kār    kardan    ādat    kard  
           Maryam    little-by-little    to work    do            habit    do.PST.3SG  
           'Maryam got used to working little by little.'
- c. derāxt-hā    zud    rošd            kard-and  
           tree-PL        early    growing    do.PST-3PL  
           'The trees grew quickly.'
- d. māhiat=e    vāqeʔi=aš        dar    raftār=aš            boruz        kard  
           nature=EZ    true=her/his    in    behavior=her/his    revelation    do.PST.3SG  
           'Her/His real nature revealed in her/his behavior.'

For sake of space, we apply the three criteria introduced above only to *gir kardan*. Their application to the other CPrs gives similar results, singling out these CPrs as clearly "non-agentive" or unaccusative.

<sup>7</sup> For more examples, see Samvelian (2012) and the *PersPred* Database (Samvelian and Faghiri 2013).

- (17) a. #mu-hā=ye maryam amdan lā=ye zip gir kard  
 hair-PL=EZ Maryam on-purpose between=EZ zip blocking do.PST.3SG  
 ‘#Maryam’s hair was caught in the zip on purpose.’
- b. \*mu-hā=ye maryam lā=ye zip in gir=rā kard  
 hair-PL=EZ Maryam between=EZ zip this blocking=DOM do.PST.3SG
- c. \*lā=ye zip gir kard-e šo-d  
 between=EZ zip blocking do.PST-PTCP become-PST.3SG

• The verb *gereftan* ‘to take’

Likewise, *gereftan* forms agentive, as in (18), as well as non-agentive or unaccusative CPrs, as in (19):

- (18) a. maryam otāq=rā andāze gereft  
 Maryam room=DOM measure take.PST.3SG  
 ‘Maryam measured the room.’
- b. maryam duš gereft  
 Maryam shower take.PST.3SG  
 ‘Maryam took a shower.’
- c. maryam tasmim gereft ke az irān be-rav-ad  
 Maryam decision take.PST.3SG that from Iran SUBJ-go.PRS-3SG  
 ‘Maryam decided to leave Iran.’
- (19) a. in kār (tavassot=e maryam) anjām gereft  
 this work by=EZ Maryam accomplishment take.PST.3SG  
 ‘This work was done by Maryam.’
- b. xāne ātaš gereft  
 house fire take.PST.3SG  
 ‘The house caught fire.’
- c. zaxm-hā eltiām gereft-and  
 wound-PL healing take.PST-3PL  
 ‘The wound healed.’

As for aspectual properties *gereftan* can form various types of CPrs:

- Dynamic CPrs: *aks gereftan* ‘to take photos’, *andāze gereftan* ‘to measure’.
- Stative CPrs: *ruze gereftan* ‘to fast’, *mātam gefertan* ‘to be sorrowful’.
- Durative CPrs: *ruze gereftan* ‘to fast’, *āftāb gereftan* ‘to sunbathe’.

- Punctual CPrs: *pāyān gereftan* 'to end', *bārān gereftan* 'to begin to rain'.

• The verb *dāštan* 'to have'

Contrary to verbs like *zadan* and *kardan*, *dāštan* 'to have' is claimed by Folli et al. to form only stative CPrs. This claim is also contradicted by the following examples:

- (20) a. maryam name=rā be omid ersāl dāšt  
Maryam letter=DOM to Omid sending have.PST.3SG  
'Maryam sent the letter to Omid.'
- b. dowlat pāyān=e jang=rā e?lām dāšt  
government end=EZ war=DOM announce have.PST.3SG  
'The government announced the end of the war.'
- c. maryam romān=aš=rā be mādar=aš taqdim dāšt  
Maryam novel=her=DOM to mother=her dedication have.PST.3SG  
'Maryam dedicated her novel to her mother.'

In all these examples, *dāštan* alternates with *kardan* without changing the meaning of the CPr. The only notable difference between the two variants is the language register, the variant with *dāštan* is used in the formal register. Note also that contrary to Folli et al.'s claim, here the eventive properties of the non-verbal element are responsible for the dynamic interpretation of the CPr.

• The verb *bastan* 'to close, to tie'

*Bastan* also can form agentive CPrs, as in (21), as well as unaccusative CPrs, as in (22).

- (21) a. maryam bā omid qarārdād bast  
Maryam with Omid agreement close.PST.3SG  
'Maryam made an agreement with Omid.'
- b. polis tazāhor-konande-gān=rā be golule bast  
police demonstrator-PL=DOM to bullet close.PST.3SG  
'The police shot at the demonstrators.'
- c. moštari-hā jelo=ye maqāze saf bast-and  
customer-PL front=EZ shop line close.PST-3PL  
'The customers lined up in front of the shop.'
- (22) a. dast-hā=yaš pine bast-e=and  
hand-PL=her/his callosity close.PST-PTCP=COP.3PL  
'Her/his hands have become callous.'

- b. divār kamkam xaze bast  
 wall little-by-little moss close.PST.3SG  
 ‘The wall was little by little covered by moss.’
- c. āb yax bast  
 water ice close.PST.3SG  
 ‘The water froze.’

From the data examined in this section, it can be concluded that the contribution of the LV is not consistent through its combination with various non-verbal elements to form a CPr neither with respect to agentivity nor to aspectual properties. A compositional approach to Persian CPrs that determines the contribution of the verb *a priori* and regardless of the other member of the CPr is thus doomed to failure, unless all the above-mentioned counterexamples are overlooked.

### 3.1.2. The Non-Consistent Contribution of the Non-Verbal Element

The contribution of the non-verbal element of the CPr is also assumed to be consistent by compositional approaches. In Folli et al.’s (2005) analysis, for instance, the non-verbal element is responsible for *aktionsart*. Adjectives and prepositional phrases form a Small Clause with their complement and thus always give rise to telic CPrs (accomplishment or achievement), while CPrs formed with a nominal element may either be telic or atelic, depending on the noun: eventive nouns form telic and non-eventive nouns atelic CPrs.

In this section we will provide examples against Folli et al.’s assumption.

#### 3.1.2.1. The Non-Consistency of Adjectives

The examples below show that CPrs containing an adjective are not systematically telic.

- (23) a. maryam xabar=rā az omid penhān dāšt  
 Maryam news=DOM from Omid hidden have.PST.3SG  
 ‘Maryam hid the news from Omid.’
- b. maryam xabar=rā se sāl az omid penhān dāšt  
 Maryam news=DOM three year from Omid hidden have.PST.3SG  
 ‘Maryam hid the news from Omid for three years.’
- c. #maryam xabar=rā dar yek sāʔat az omid penhān dāšt  
 Maryam news=DOM in one hour from Omid hidden have.PST.3SG  
 (Lit.) ‘#Maryam hid the news from Omid in one hour.’

- (24) a. *maryam omid=rā az daqdaqe dur dāšt*  
 Maryam Omid=DOM from worry far have.PST.3SG  
 ‘Maryam kept Omid far from worries.’
- b. *maryam omid=rā sāl-hā az daqdaqe dur dāšt*  
 Maryam Omid=DOM year-PL from worry far have.PST.3SG  
 ‘Maryam kept Omid far from worries for years.’
- c. *#maryam omid=rā dar panj daqiqe az daqdaqe dur dāšt*  
 Maryam Omid=DOM in five minute from worry far have.PST.3SG  
 (Lit.) ‘#Maryam kept Omid far from worries in five minutes.’
- (25) a. *maryam pul lāzem dāšt*  
 Maryam money necessary have.PST.3SG  
 ‘Maryam needed money.’
- b. *maryam sāl-hā pul lāzem dāšt*  
 Maryam year-PL money necessary have.PST.3SG  
 ‘Maryam needed money for years.’
- c. *#maryam dar panj daqiqe pul lāzem dāšt*  
 Maryam in five minute money necessary have.PST.3SG  
 (Lit.) ‘#Maryam needed money in five minutes.’

The CPRs used in these examples, namely *penhān dāštan* ‘to keep hidden’ (Lit. ‘hidden have’), *dur dāštan* ‘to keep far away’ (Lit. ‘far have’) and *lāzem dāštan* ‘to need’ (Lit. ‘necessary have’) are formed with an adjective and nevertheless are not telic.

### 3.1.2.2. The Non-Consistency of Prepositional Phrases

Like adjectives, PPs are claimed by Folli et al. (2005) to form only telic predicates. Once again, the existence of several atelic CPRs containing a PP goes against this generalization. In the following examples, *be masxare gereftan* ‘to mock’ (Lit. ‘to mocking take’) and *be ohde dāštan* ‘to have the responsibility of, to be in charge of’ are two examples of atelic CPRs formed with a PP.

- (26) a. *maryam omid=rā be masxare gereft*  
 Maryam Omid=DOM to mocking take.PST.3SG  
 ‘Maryam mocked Omid.’
- b. *maryam omid=rā sāʔat-hā be masxare gereft*  
 Maryam Omid=DOM hour-PL to mocking take.PST.3SG  
 ‘Maryam mocked Omid for hours.’

c. #maryam omid=rā dar yek sāʔat be masxare gereft  
 Maryam Omid=DOM in one hour to mocking take.PST.3SG  
 (Lit.) ‘#Maryam mocked Omid in one hour.’

(27) a. maryam esteqbāl az dānešju-yān=rā be ohde dār-ad  
 Maryam welcome of student-PL=DOM to charge have.PRS-3SG  
 ‘Maryam is in charge of welcoming the students.’

b. maryam sāl-hā esteqbāl az dānešju-yān=rā be ohde dāšt  
 Maryam year-PL welcome of student-PL=DOM to charge have.PST.3SG  
 ‘Maryam was in charge of welcoming the students for years.’

c. #maryam dar yek sāʔat esteqbāl az dānešju-yān=rā be ohde dāšt  
 Maryam in one hour welcome of student-PL=DOM to charge have.PST.3SG  
 (Lit.) ‘#Maryam was in charge of welcoming the students in one hour.’

### 3.1.2.3. *The Non-Consistency of Nouns*

The following examples show that non-eventive nouns, contrary to Folli et al’s (2005) claim, can form telic CPRs:

(28) a. maryam otāq=rā andāze gereft  
 Maryam room=DOM measure take.PST.3SG  
 ‘Maryam measured the room.’

b. #maryam otāq=rā panj daqiqe andāze gereft  
 Maryam room=DOM five minute measure take.PST.3SG  
 (Lit.) ‘#Maryam measured the room for five minutes.’

c. maryam otāq=rā dar čand sāniye andāze gereft  
 Maryam room=DOM in few second measure take.PST.3SG  
 ‘Maryam measured the room in a few seconds

(29) a. nān kapak zad  
 bread mold hit.PST.3SG  
 ‘This bread went moldy.’

b. #nān se ruz kapak zad  
 bread three day mold hit.PST.3SG  
 (Lit.) ‘#This bread went moldy for three days.’

c. nān ye-ruze kapak zad  
 bread in-one-day mold hit.PST.3SG  
 ‘This bread went moldy in one day.’

- (30) a. *mār pust andāxt*  
 snake skin throw.PST.3SG  
 ‘The snake sloughed off.’
- b. *#mār do ruz pust andāxt*  
 snake two day skin throw.PST.3SG  
 (Lit.) ‘#The snake sloughed off for two days.’
- c. *mār ye-ruze pust andāxt*  
 snake in-one-day skin throw.PST.3SG  
 ‘The snake sloughed off in one day.’

To conclude, almost *none* of the empirical generalizations on the respective contribution of the verb and the non-verbal element stated by Folli et al. (2005) hold when a larger range of data is taken into account.

### 3.2. The Non-Predictable Semantic Content of the CPr

The non-predictability of the meaning of the CPr is another significant impediment to fully compositional approaches. In order for the latter to work, the meaning of the CPr must be derivable on the basis of the meaning of its components. However, as mentioned in several studies (Goldberg 1996; Karimi-Doostan 1997; Family 2006; Bonami and Samvelian 2010; Samvelian 2012; Samvelian and Faghiri 2013; among others), numerous Persian CPrs are semantically opaque. Moreover, as shown by Samvelian (2012) and Bonami and Samvelian (2010), even when a CPr is semantically transparent, it is barely ever the case that its meaning is fully predictable from the meaning of its component parts. CPrs are subject to various levels of lexicalization, in a way that closely parallels what is seen with lexemes formed by morphological means.<sup>8</sup>

#### 3.2.1. Specialization

In some cases, the meaning of the CPr is a specialization of the predictable meaning of the combination: *čāqu zadan* (Lit. ‘knife hit’), is not only to hit somebody with a knife but to stab them; *dast dādan* (Lit. ‘hand give’) does not only imply that you put your hand in somebody’s else hand, but that you shake hands with them; *āb dādan* (Lit. ‘water give’), like *to water* in English, is not just pouring water on something, and finally *šir dādan* (Lit. ‘milk give’) is not just the action of giving milk to somebody, but to breast-feed them.

<sup>8</sup> There are of course instances of CPrs with a fully predictable meaning. These are mostly CPrs formed by a predicative noun, which conveys the entire lexical meaning of the CPr, e.g. *anjām dādan* ‘to realize’ (Lit. ‘accomplishment give’), *tekān xordan* ‘to move’ (Lit. ‘movement collide’), *faryād zadan* ‘to shout, to cry’ (Lit. ‘shout hit’) and *entexāb kardan* ‘to choose’ (Lit. ‘choice do’).

### 3.2.2. Semantic Drift

In other examples semantic drift has taken place, either by metaphor or by metonymy. The link between the compositional meaning and the lexicalized meaning is sometimes still recoverable synchronically. For instance, the lexicalized meaning of *guš kardan* ‘to listen’ (Lit. ‘ear do’) can be recovered via metonymy: The CPr designates the prototypical action accomplished by ears. In *zanjir zadan*, the elliptical element of the meaning, *sine* ‘chest’, can also be recovered: the CPr comes from *bā zanjir (be) sine zadan* ‘to hit one’s chest with chains’ or *zanjir be sine zadan* ‘to hit chains against one’s chest’. However, in numerous other cases the initial link is no longer perceivable by the speakers. *ru gereftan* ‘to become cheeky’ (Lit. ‘face take’) and *dast andāxtan* ‘to mock’ (Lit. ‘hand throw’), for instance constitute opaque sequences in synchrony.

Folli et al. (2005) do not propose any explicit treatment for the issue raised by these examples, namely the fact that argument structure and eventive properties does not exhaust the meaning of a CPr and even if one assumes that they can be determined on a constructionist basis, still they do not allow to determine the exact meaning of most of CPrs, which can be more or less idiomatic.

### 3.3. The Collocational Association between the CPr Components

Another problem faced by fully compositional accounts is the collocational association between the verb and the non-verbal element of the CPr. Thus, it may happen that two synonymous nouns combine with two different verbs to form synonymous CPrs, as it is the case with *rašk* and *hesādat*, both meaning ‘envy’ or ‘jealousy’, which respectively combine with *bordan* ‘to take’ and *kardan* ‘to do’ to form *rašk bordan* and *hesādat kardan*, both meaning ‘to envy’ or ‘to be jealous’. Likewise, *āroq* ‘burp’ combines with *zadan* ‘to hit’, *āroq zadan* ‘to burp’, while *sorfe* ‘cough’, which is semantically close to *āroq*, combines with *kardan*, *sorfe kardan* ‘to cough’. However, such facts have either been overlooked or minimized in the studies on Persian CPrs. Folli et al. (2005), for instance, acknowledge this collocational association but do not address it thoroughly and state hasty generalizations on the basis of a restricted set of examples:

- (31) a. *miz=o tamiz kard-am* (Folli et al. 2005:1389, (60))  
       *table=DOM clean do.PST-1SG*  
       ‘I cleaned the table.’
- b. *miz tamiz šo-d*  
       *table clean become-PST.3SG*  
       ‘The table got/became cleaned.’

- (32) a. *bačča=ro hamum kard-am* (Folli et al. 2005:1389, (62))  
 child=DOM bath do.PST-1SG  
 ‘I bathed the child.’  
 b. \**bačče hamum šo-d*  
 child bath become-PST.3SG  
 (intended) ‘The child was given a bath.’

These examples show that while *tamiz* ‘clean’ and *hamum* ‘bath’ both combine with *kardan* ‘to do’ to form agentive CPrs, only *tamiz* ‘clean’ combines with *šodan* ‘to become’ to form an unaccusative CPr. According to Folli et al. (2005:1389), such restrictions are not idiosyncrasies and can be accounted for on the basis of the syntactic properties of *šodan*, which “selects for a predicative small clause complement, while *kardan* can select for either a nominal complement (when it gets a ‘do’ meaning) or a small clause complement (when it gets a ‘make’ reading).” This explains the success of a *kardan/šodan* alternation in CPrs with a predicative non-verbal element, and its failure in CPrs with a nominal element. Let us now consider the following examples:

- (33) a. *polis maryam=rā jarime kard*  
 police Maryam=DOM fine do.PST.3SG  
 ‘The policeman fined Maryam.’

- b. *maryam jarime šo-d*  
 Maryam fine become-PST.3SG  
 ‘Maryam was fined.’

- (34) a. *divār-hā=rā pārsāl rang kard-im*  
 wall-PL=DOM last-year paint do.PST-1PL  
 ‘[We] painted the walls last year.’

- b. *divār-hā=rā pārsāl rang šo-d-and*  
 wall-PL=DOM last-year paint become-PST-3PL  
 ‘The walls were painted last year.’

- (35) a. *maryam dar=rā qofl kard*  
 Maryam door=DOM lock do.PST.3SG  
 ‘Maryam locked the door.’

- b. *dar qofl šo-d*  
 door lock become-PST.3SG  
 ‘The door was locked.’

These examples show that there is no ban on the combination of nominal non-predicative nouns with *šodan* to form unaccusative CPrs, which implies in turn that the restrictions exemplified by Folli et al. in examples (31) and (32) cannot receive a syntactic explanation, but are idiosyncratic in nature.

The same problem rises for some other restrictions noted by Folli et al., namely *dād zadan* ‘to shout’ (Lit. ‘shout hit’) vs. \**dād xordan* (Lit. ‘shout collide’); *kār kardan* ‘to work’ (Lit. ‘work do’) vs. \**kār xordan* (Lit. ‘work collide’). Here is the explanation provided by the authors:

“(…) certain NV elements may not be combined with the unaccusative *xordan* ‘collide’, because the events that they denote can only be caused agentively – they are, in essence, inherently unergative.” (Folli et al. 2005:1390)

No criterion is provided for the identification of inherently unergative nouns, but let us admit that *dād* and *kār* are indeed such nouns. We note however that although *kār* does not combine with *xordan* to form an unaccusative CPr, it nevertheless combines with *šodan* ‘to become’:

(36) *darbāre=ye in matlab besyār kār šo-d-e ast*  
 about=EZ this subject very work become-PST-PTCP COP.3SG  
 ‘A great deal of work has been done on this topic.’

Moreover, other “inherently unergative” nouns like *sili* ‘slap’ or *kotak* ‘beating’ combine also with *xordan* to form unaccusative CPrs:

(37) *omid az sārā sili xor-d*  
 Omid from Sara slap collide-pst.3sg  
 ‘Omid was slapped by Sara.’

Consequently, the fact that *dād* does not combine with *xordan* cannot be accounted for on the basis of its syntactic or semantic properties. This restriction is idiosyncratic and must be stipulated.

To conclude, in this section we showed that *none* of the generalizations stated by Folli et al. (2005) withstand a close scrutiny.

#### 4. Compositionality Revisited

Despite all the problems mentioned in the previous section, we claim that a compositional account of Persian CPrs is still possible, provided compositionality is

defined *a posteriori*, like in idiomatically combining expressions, in the sense of Nunberg *et al.* (1994). This view of Persian CPrs can be developed into a Construction-based approach (Goldberg 1995), where:

- 1) Each CPr corresponds to a Construction.
- 2) CPrs can be grouped in classes according to their semantic and syntactic properties and each class can be represented by a partially fixed Construction.
- 3) Constructions can be structured in networks, thus accounting for different semantic and syntactic relations between CPrs.

#### 4.1. Persian CPrs as Idiomatically Combining Expressions

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 its meaning can be distributed among 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. This is the line of argumentation developed by Nunberg *et al.* (1994) to support a compositional view of expressions such as *spill the beans*. Although the meaning of the sequence is conventional or idiomatic, in that it must be learned and cannot be driven from its component parts, the sequence is still compositional, in that its meaning can be distributed among its components. In the context of its association with *beans*, *spill* means ‘to divulge’, and in the context of its association with *spill*, *bean* means ‘secret’. This view of Persian CPrs is defended in recent studies by Müller (2010), Samvelian (2012) and Samvelian and Faghiri (*in press*). Table 2 illustrates this point for a set of CPrs formed with *zadan*. Each line contains examples of CPrs where the verb can be assigned a meaning comparable to that of a lexical verb in English, in combination with a set of nouns.

An important advantage of this approach is that not only it provides a fine-grained and empirically valid compositional analysis of Persian CPrs, but it also allows to account for their productivity. 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 *barčasb zadan* ‘to label’ (Lit. ‘label hit’), *tambr zadan* ‘to stamp’ (Lit. ‘stamp hit’), etc.

Samvelian (2012) provides a thorough study of CPrs formed with *zadan* ‘to hit’. On the basis of a detailed syntactic and semantic classification, the author establishes a

list of abstract Constructions, each associated to a syntactic structure, argument structure and an abstract meaning.

Table 2. Meanings of *zadan* in the context of its CPRs

CPRs formed with <i>zadan</i> (N + <i>zadan</i> )	The meaning of <i>zadan</i>
<i>kare zadan</i> 'to butter', <i>lāk zadan</i> 'to varnish', <i>rang zadan</i> 'to paint', <i>šāmpu zadan</i> 'to shampoo', <i>vāks zadan</i> 'to polish', ...	'to apply'
<i>adviye zadan</i> 'to put spice', <i>fefel zadan</i> 'to pepper', <i>namak zadan</i> 'to salt', ...	'to add', 'to incorporate'
<i>barčāsb zadan</i> 'to label', <i>dastband zadan</i> 'to handcuff', <i>lejām zadan</i> 'to bridle', <i>mangane zadan</i> 'to staple', <i>tambr zadan</i> 'to stamp', ...	'to put'
<i>eynak zadan</i> 'to wear glasses', <i>kerāvāt zadan</i> 'to wear a tie', <i>māsk zadan</i> 'to wear a mask', ...	'to wear'
<i>javāne zadan</i> 'to bud', <i>juš zadan</i> 'to sprout', <i>kapak zadan</i> 'to go mouldy', <i>šabnam zadan</i> 'to dew', <i>šokufe zadan</i> 'to bloom', <i>tabxāl zadan</i> 'to develop coldsore', <i>tāval zadan</i> 'to blister', <i>zang zadan</i> 'to rust', <i>pine zadan</i> 'to become calloused', ...	'to develop', 'to form'
<i>pol zadan</i> 'to build a bridge', <i>jādde zadan</i> 'to build a road', <i>sad zadan</i> 'to build a dam', <i>saqf zadan</i> 'to build a ceiling', ...	'to build'
<i>čādor zadan</i> 'to set up a tent', <i>dārbast zadan</i> 'to erect a scaffolding', <i>ordu zadan</i> 'to set a camp', <i>pašeband zadan</i> 'to put up a fly sheet', ...	'to set up'
<i>arbade zadan</i> 'to yell', <i>dād zadan</i> 'to shout', <i>faryād zadan</i> 'to shout', <i>jiq zadan</i> 'to scream', ...	'to emit (a cry)'
<i>ar zadan</i> 'to bray', <i>čahčahe zadan</i> 'to sing (bird)', <i>jik zadan</i> 'to chirp', <i>šeyhe zadan</i> 'to neigh', ...	'to emit (an animal cry)'

#### 4.2. Persian CPRs as Constructions

A Construction, in the sense of Goldberg (1995) and Kay and Fillmore (1999), is a conventional association between a form and a meaning. Given that Persian CPRs have a conventional meaning, they each correspond to a Construction, and are thus comparable to lexemes. Constructions can be of various levels of abstractness and can be organized hierarchically, going from the most specific ones to the more abstract ones. Samvelian (2012) applies this approach to a set of CPRs formed with the verb *zadan* 'to hit' and a nominal element. The latter are grouped in semantic classes, with various degrees of coherence. Each class corresponds to a partially fixed Construction. Here are a few examples of these Constructions:

(38) Spreading-*zadan* Construction

N0 (be)N1 N2 V

Agent Ground Figure

'N0 applies N2 on N1' or 'N0 covers the surface of N1 with N2'

(39) Instrument-*zadan* Construction

N0 (be) N1 N2 V

Agent Patient Instrument

'N0 accomplishes the typical action for which N2 is used (on N1)'

(40) Slandering-*zadan* Construction

N0 be N1 N2 V

Slanderer Slanderee Slander

'N0 accuses N1 of N2'

(41) Forming-*zadan* Construction

N0 N1 V

Location Theme

'N1 is formed on N0'

In each Construction notation:

(a) The first line gives the label of the Construction. The fact that *zadan* is specified in the label implies that the Construction is partially fixed and the verbal lexeme in the construction is one of the forms of *zadan*.

(b) The second line gives the valency (subcategorization frame) and the syntactic construction of the minimal sentence including the CP. The formalism is inspired by Gross's (1975) Lexicon-Grammar notation: N0 stands for the subject, N1 for the first complement (object), N2 for the second complement (object), etc.

(c) The third line provides the mapping between the semantic roles and the grammatical functions.

(d) The last line indicates the abstract meaning assigned to the Construction as a whole.

Here are examples of CPRs associated with each Construction:

- Spreading-*zadan* Cons: *āb* – 'to wet', *āhār* – 'to starch', *kare* – 'to butter', *rang* – 'to paint', *vāks* – 'to polish'...

- Instrument-*zadan* Cons: *jāru* – ‘to broom’, *mesvāk* – ‘to brush one’s teeth’, *otu* – ‘to iron’, *šāne* – ‘to comb’...
- Forming-*zadan* Cons: *javāne* – ‘to blossom’, *juš* – ‘to sprout’, *kapak* – ‘to mouldy’, *tāval* – ‘to blister’, *zang* – ‘to go rusty’...
- Slandering-*zadan* Cons: *ang* –, *bohtān* –, *ettehām* –, *tohmat* – ‘to slander’...

Note that in this approach, all lexicalized CPRs are listed in the lexicon, contrary to Folli et al.’s approach, which claims that Persian CPRs, despite their idiomaticity, need not to be listed.<sup>9</sup>

### 4.3. Constructions in Networks

We adopt furthermore a *type hierarchy* or an *inheritance hierarchy* of Constructions of varying levels of specificity (Michaelis, to appear). Constructions are thus structured in networks, which represent different relationships such as hyponymy/hyperonymy (subtypes vs. supertypes), synonymy and valency alternations. Each Construction belongs to several different Construction types. Consequently, a given Construction type can inherit properties from multiple dominating types.

#### 4.3.1. Subclasses and Superclasses

Some semantic classes can be grouped together into a more abstract class. In this case, the Construction associated with them is the subtype of a less specific Construction. For instance, the Spreading-*zadan* Construction can be considered as a subtype of Locatum Construction. Locatum verbs (Clark and Clark 1979), e.g. *to paint*, *to salt*, incorporate a Figure (i.e. the noun to which the verb is morphologically related) and have a Ground argument realized as an NP or a PP: ‘to paint sth’ = ‘to put paint (Figure) on sth (Ground)’. As for Persian Locatum CPRs, the Figure is the nominal element of the CPR:

#### (42) Locatum-*zadan* Construction

N0	(be) N1	N2	V
Agent	Ground	Figure	
‘N0 puts/incorporates N2 on/into N1’			

---

<sup>9</sup> Note however that this does not imply that the whole sequence is listed as a lexical entry in the lexicon. For an implementation of this approach see Müller (2010).

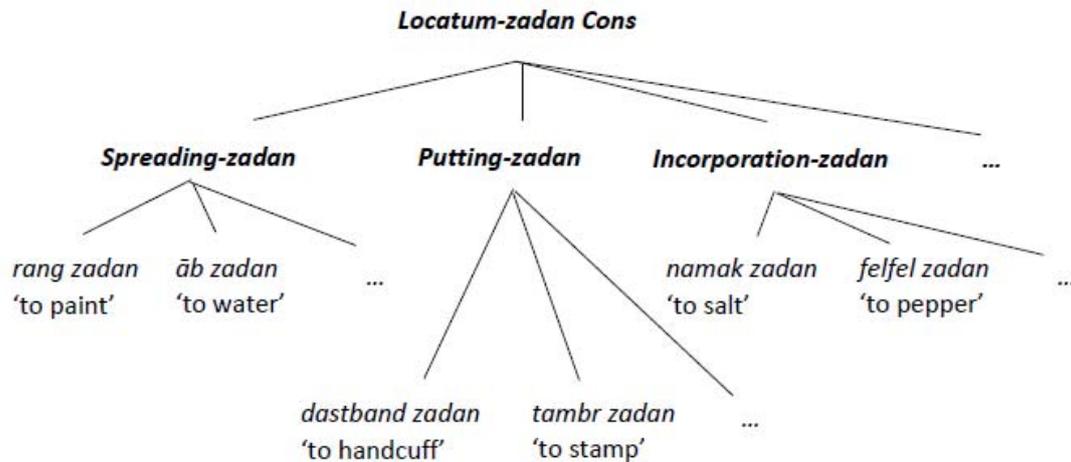


Figure 1. Subtypes of Locatum-zadan Construction

The subtypes of the Locatum-zadan Construction, that is, the Spreading-zadan Construction, the Incorporation-zadan Construction and the Putting-zadan Construction (see Figure 1), all have an Agent and a Ground argument; the nominal element of the CP<sub>r</sub> is the Figure. They diverge in the manner that the Figure is placed on or inside the Ground argument. In the predicates realizing the Spreading-zadan Construction, the Figure is spread or applied on the surface of the Ground. The verb *zadan* in these CP<sub>r</sub>s can generally alternate with the lexical verb *mālidan* 'to apply', 'to spread'. In the CP<sub>r</sub>s corresponding to the Incorporation-zadan Construction, the Figure is incorporated into the ground. The verb *zadan* is then synonymous to *rixtan* 'to pour'. Finally, in Putting-zadan Construction, the Figure is placed on the Ground. These differences may entail further ontological distinctions. For instance, at the end of the event denoted by a CP<sub>r</sub>, which is an instance of the Incorporation-zadan Construction, the Figure can be confounded with the Ground or absorbed by it.

#### 4.3.2. *Synonymous Constructions*

The same Construction can be realized by different verbs, e.g. *kardan* 'to do' and *kešidan* 'to pull' also form Instrumental predicates, e.g. *jāru kardan* and *jāru kešidan* 'to broom'. So, along with the Instrument-zadan Construction, there are also the Instrument-kešidan Construction and the Instrument-kardan Construction. These three partially fixed Constructions are sub-types of a more abstract Construction, with no lexically fixed element, that is, the Instrument Construction. Synonymy rises when the same noun occurs in the same Construction realized by different verbs.

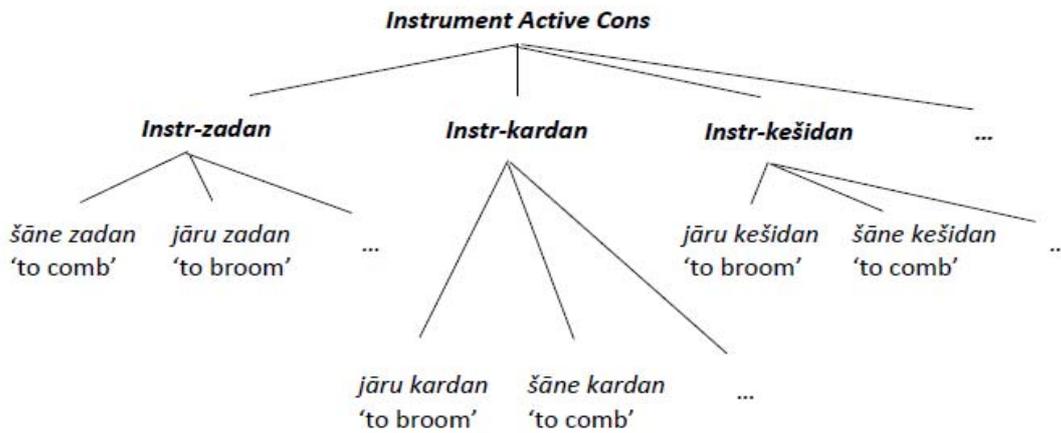


Figure 2. Instrument (active) Construction

4.3.3. Valency Alternating Constructions

The same Construction can display valency alternations. For instance, in an Instrument Construction, the Agent can be mapped to the grammatical subject and the Patient to the grammatical object, which gives rise to an “active” Instrument Construction, or the Patient can be mapped to the grammatical subject, forming a “passive” or “unaccusative” Instrument Construction. Valency alternations in CPRs are often realized by the choice of the verb: *otu zadan* ‘to iron’ vs. *otu xordan* ‘to be ironed’ (Lit. ‘iron collide’), *ātaš zadan* ‘to set fire’ vs. *ātaš gereftan* ‘to take fire’ (Lit. ‘fire take’).

All of the relations described above can be structured in a network using multiple inheritance hierarchies. Figure 3 provides a partial hierarchy of Persian CPRs.

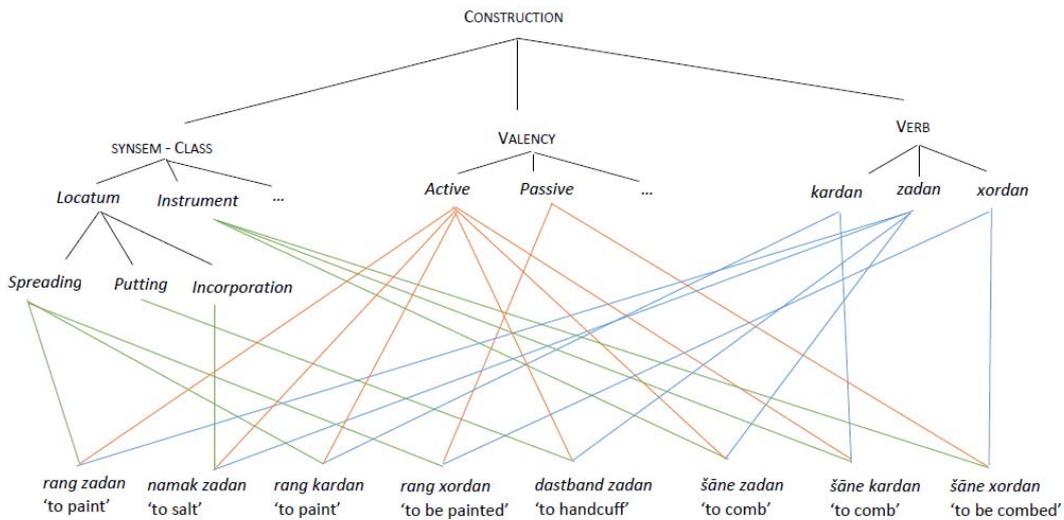


Figure 3. Partial hierarchy of Persian CPRs

#### 4.4. Class Coherence, Size and Productivity

In our 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 (see Table 2). However, we want to defend the idea that productivity is not always compositionality-based and that non-compositional Constructions (or classes) can also be productive. The Communicating-*zadan* Construction, e.g. *telefon zadan* ‘to phone’ (Lit. ‘phone hit’), *telegrāf zadan* ‘to send a telegraph’ (Lit. ‘telegraph hit’), *bisim zadan* ‘to walkie-talkie’, ‘to communicate by means of a walkie-talkie’ (Lit. ‘walkie-talkie hit’), illustrates this situation. While it is impossible to assign a meaning to *zadan* in these combinations, recent combinations such as *imeyl zadan* ‘to email’ or *esemes zadan* ‘to text, to sms’ are nevertheless created by analogical extension. We furthermore hypothesize that the productivity of Persian CPrs is also related to other parameters such as the coherence of the classes and their size.

The Construction associated to each class is an abstraction based on the most salient and regularly shared properties of its members. Classes vary in their coherence and the meaning associated to the Construction can be more or less abstract or specific. In some classes, the meaning of each CPr can be straightforwardly derived from the meaning associated to the Construction. This is the case in the Spreading-*zadan* Construction: For each member the paraphrase ‘N0 applies N2 on N1’ provides its meaning and its syntactic properties. This situation yields a fully compositional class, where the verb is comparable to a semi-lexical or lexical verb and the relation between the verb and the nominal element is comparable to the semantic selection of an argument by a verb. One could even consider the creation of a lexical entry for the verb *zadan* with the meaning of ‘to apply’, especially with respect to the fact that *zadan* alternates with *mālidan* ‘to apply’ in these combinations. Consequently, the class is highly productive and listing all potential combinations, apart from lexicalized ones, is impossible.

Some other groupings are based on more abstract properties. For instance, CPrs corresponding to the Forming Construction denote heterogeneous processes implying the eruption, the forming or the transformation of an entity, e.g. *kapak zadan* ‘to go mouldy’ (Lit. ‘mould hit’), *tāval zadan* ‘to blister’ (Lit. ‘blister hit’), *yax zadan* ‘to freeze’ (Lit. ‘ice hit’), *zang zadan* ‘to go rusty’ (Lit. ‘rust hit’). Contrary to the previous case, the nominal element is not exclusively selected via its conceptual properties and the link between the components is more collocational: *javāne zadan* ‘to blossom’ (Lit. ‘blossom hit’) vs. *gol dādan* ‘to flower’ (Lit. ‘flower give’) and not *gol zadan*. Consequently, one

can assume that the productivity of this class is not comparable to that of the previous one.

Classes also vary with respect to their size. It is expected that fully compositional classes should be large, since new combinations can regularly enrich the class. This is indeed the case with the Spreading-*zadan* Construction, for instance, which constitutes one of the largest classes of CPRs formed with *zadan*. However, some compositional classes are restricted due to the small number of nouns that could fit in the Construction. For instance, the Scream Emission Construction corresponds to only about twenty CPRs, which is not surprising, given the number of nouns denoting a cry in Persian.

Finally, the issue of the productivity cannot be investigated without integrating the concurrent possibilities, that is, the competing verbs to form a CPR: various verbs can be used with the same non-verbal element to realize the same “verbal concept”, thus giving rise to the existence of synonymous CPRs. This situation is reminiscent of the choice between competing affixes to form a new word in construction morphology.

## 5. Conclusion

The “fully” compositional view of Persian have become so prevailing that it now seems to be taken for granted. In this paper we advocated for a view going against this widespread consensus. We showed that not only the lexical meaning of Persian CPRs is barely ever fully predictable from the meaning of their component parts, but also that even more abstract properties, such as argument and event structure, cannot be determined *a priori*, on the basis of solely one component of the CPR regardless of the other one.

In line with Goldberg (1996) and Family (2006), we argued that Persian CPRs must be listed since they correspond to conventional pairings of forms and meanings. However, we defended the idea that there is no contradiction between storage and a compositional account of these combinations, provided compositionality is defined *a posteriori*, in the sense of Nunberg et al. (1994) for idiomatically combining expressions. The Construction-based approach we outlined is based on the assumption that despite their idiomacity, Persian CPRs can be grouped upon their syntactic and semantic similarities. Productivity results in this account from the possibility to analogically extend the existing classes and can be compositionality-based or not.

To finish, we want to insist on the importance of considering a large set of data before drawing generalizations on CPR formation in Persian. Like morphological lexeme formation processes, CPR formation in Persian displays a great amount of

idiosyncrasies. Regularities apply within small and medium-sized sets and rarely concern the whole set of combinations formed with a given verb. A careful investigation of data is necessary if sound and reliable generalizations are intended.

## References

- Bashiri, Iraj. 1981. *Persian Syntax*. Minneapolis: Burgess Publishing Company.
- Bonami, Olivier, and Pollet Samvelian. 2010. Persian complex predicates: Lexeme formation by itself. Paper presented at Septièmes Decembrettes Morphology Conference, Toulouse.
- Bonami, Olivier, and Pollet Samvelian. in press. The diversity of inflectional periphrasis in Persian. *Journal of Linguistics*.
- Borer, Hagit. 1994. The projection of arguments. In *University of Massachusetts Occasional Papers in Linguistics 17, Syntax and Semantics*, ed. by Elena Benedicto and Jeff Runner. GLSA, University of Massachusetts, Amherst.
- Butt, Miriam. 1995. *The structure of complex predicates*. Stanford: CSLI Publications.
- Clark, Eve V., and Herbert H. Clark. 1979. When nouns surface as verbs. *Language* 55(4):767–811.
- Dabir-Moghaddam, Mohammad. 1982. Passive in Persian. *Studies in the Linguistic Sciences* 12(1):63–90.
- Family, Niloufar. 2006. *Explorations of semantic space: The case of light verb constructions in Persian*. Doctoral dissertation, EHESS.
- Folli, Raffaella, Heidi Harley, and Simin Karimi. 2005. Determinants of event type in Persian complex predicates. *Lingua* 115:1365–1401.
- Ghomeshi, Jila. 1997. Topics in Persian VPs. *Lingua* 102:133–167.
- Ghomeshi, Jila, and Diane Massam. 1994. Lexical/syntactic relations without projections. *Linguistic Analysis* 23(3/4):75–217.
- Goldberg, Adele E. 1995. *A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 1996. Words by default: Optimizing constraints and the Persian complex predicate. In *Annual Proceedings of the Berkeley Linguistic Society* 22, 132–146. Berkeley.
- Grimshaw, Jane, and Armin Mester. 1988. Light verbs and theta-marking. *Linguistic Inquiry* 19:205–232.
- Gross, Maurice. 1975. *Méthodes en syntaxe: Régime des constructions complétives*. Paris: Hermann.
- Hale, Kenneth L., and Samuel J. Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In *The View from Building 20*, ed. by Kenneth L. Hale and Samuel J. Keyser, 53–109. Cambridge, MA: MIT Press.
- Hale, Kenneth L., and Samuel J. Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.
- Karimi, Simin. 1996. Case and specificity: Persian *râ* revisited. *Linguistic Analysis* 26:74–194.
- Karimi-Doostan, Gholamhossein. 1997. Light verb constructions in Persian. Doctoral dissertation, University of Essex.
- Kay, Paul, and Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. *Language* 75:1–33.
- Kratzer, Angelika. 1996. Severing the external argument from its verb. In *Phrase structure and the lexicon*, ed. by Johan Rooryck and Laurie Zaring, 109–137. Dordrecht: Kluwer Academic Publishers.

- Lazard, Gilbert. 1982. Le morphème *rā* en persan et les relations actancielles. *Bulletin de la Société de Linguistique de Paris* 77(1):177–207.
- Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2):201–225.
- Megerdooimian, Karine. 2001. Event structure and complex predicates in Persian. *Canadian Journal of Linguistics* 46(1/2):97–125.
- Megerdooimian, Karine. 2012. The status of the nominal in Persian complex predicates. *Natural Language and Linguistic Theory* 30(1):179–216.
- Michaelis, Laura A. to appear. Sign-based Construction Grammar. In *The Oxford Handbook of Linguistic Analysis*, ed. by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press.
- Moyne, John A. 1974. The so-called passive in Persian. *Foundations of Language* 12:249–67.
- Müller, Stefan. 2010. Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* 46(2):601–655.
- Nunberg, Geoffrey, Ivan A. Sag, and Thomas Wasow. 1994. Idioms. *Language* 70:491–538.
- Samiian, Vida. 1983. Origins of phrasal categories in Persian: An X-bar analysis. Doctoral dissertation, UCLA.
- Samvelian, Pollet. 2001. Le statut syntaxique des objets nus en persan. *Bulletin de la Société de Linguistique de Paris* 96(1):349–388.
- Samvelian, Pollet. 2006–2007. Le sujet, l'objet et l'inaccusativité dans les prédicats complexes nom-verbe en persan. *Cahiers de Linguistique de l'INALCO* 6:155–190.
- Samvelian, Pollet. 2012. *Grammaire des prédicats complexes. Les constructions nom-verbe*. Lavoisier.
- Samvelian, Pollet, and Pegah Faghiri. 2013. Introducing PersPred, a syntactic and semantic database for Persian complex predicates. In *Proceedings of the Workshop on Multiword Expressions*. Atlanta, Georgia, USA.
- Samvelian, Pollet, and Pegah Faghiri. in press. Rethinking compositionality in Persian complex predicates. In *Proceedings of the 39th Annual Meeting of the Berkeley Linguistics Society*.
- Samvelian, Pollet, Pegah Faghiri, and Sarra El Ayari. 2014. Extending the coverage of a MWE database for Persian CPs exploiting valency alternations. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 4023–4026.
- Vendler, Zenon. 1957. Verbs and times. *Philosophical Review* 66:143–160.

**To cite this article:**

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