

Re-thinking Compositionality in Persian Complex Predicates*

Pollet Samvelian & Pegah Faghiri

Université Sorbonne Nouvelle

Introduction

Persian Complex Predicates (CPs) have been a focus of interest during the last two decades. Their formation (i.e. morphological/lexical vs. phrasal/syntactic) and their interpretation (compositional vs. idiomatic) have been thoroughly investigated and various analyses have been proposed to account for their seemingly contradictory properties. In this paper, we revisit the issue of the compositionality of Persian CPs, focusing on Noun-Verb combinations, and provide a Construction-based approach which allows to overcome the apparent contradiction between their idiomatic and compositional properties, and offers a new insight into their productivity.

The number of simplex verbs in Persian is limited to around 250, only half of which are currently used by the speech community¹. The verbal lexicon is thus mainly formed of syntactic combinations, including a verb and a non-verbal element, a noun, e.g. *qadam zadan* ‘to walk’ (Lit. ‘step hit’), an adjective, e.g. *derāz kešidan* ‘to lay down’ (Lit. ‘long pull’), a particle, e.g. *bar dāštan* ‘to take’ (Lit. ‘PARTICLE have’), or a prepositional phrase, e.g. *be kār bordan* ‘to use’ (Lit. ‘to work take’). These combinations are generally referred to as Complex Predicates

*This work was supported by the bilateral project *PerGram*, funded by the ANR (France) and the DGfS (Germany) [grant no. MU 2822/3-I] and is related to the work package LR4.1 of the Labex EFL (funded by the ANR/CGI). We would like to thank Gwendoline Fox for her helpful comments.

¹ Sadeghi (1993)’s estimation is 252 (115 commonly used) and Khanlari (1986) provides a list of 279 simplex verbs.

(CPs), Compound Verbs or Light Verb Constructions (LVCs). New “verbal concepts” are regularly coined as CPs rather than simplex verbs, e.g. *yonize kardan* ‘to ionize’ (Lit. ‘ionized do’) instead of *yonidan*. Although morphological lexeme formation process outputting verbs from nouns, e.g. *xāb* ‘sleep’ > *xāb-idan* ‘to sleep’ is available, it has ceased to be productive and is used only in a marginal way.

Their productivity, their syntactic formation, along with a certain degree of semantic transparency have favored a compositional view of Persian CPs. Although their idiomatic properties have been generally acknowledged, they have nevertheless been overlooked or minored by the studies adopting a compositional approach. The latter almost uncontroversially admit that at least the argument structure and the eventive properties, if not the lexical meaning, of these combinations can be compositionally derived, on the basis of a consistent contribution of the verb and the non-verbal element. In this paper, we argue against this consensus. After briefly introducing the generalizations stated by the most recent compositional studies, we show their empirical inconsistency and conclude that Persian CPs are multiword expressions with a conventional meaning, and thus need to be stored. We furthermore claim that a compositional approach is nevertheless possible provided compositionality is defined *a posteriori*, in the sense of Nunberg *et al.* (1994). We then outline a Construction-based approach illustrating these points.

1. Previous Compositional Approaches

Two main arguments have been invoked in favor of a compositional analysis of Persian CPs: a) The predictability of their argument and event structure; b) The predictability of their lexical (referential) meaning.

In the examples below, the referential meaning of the CP and the roles assigned to the event participants is determined by the nominal element, since the semantic participants of the CP, ex. (1b), are identical to those of the noun within the NP it projects, ex. (1a). The verb on the other hand determines the argument mapping, since the substitution of the verb by another one entails a change in the mapping between the participants and the grammatical functions, ex. (1b) and (1c).

- | | | |
|-----|---|--|
| (1) | <p>a. <i>sili=e Sārā be Omid</i>
slap=EZ Sara to Omid
‘Sara’s slap to Omid’²</p> <p>b. <i>Sārā be Omid sili zad</i>
Sara to Omid slap hit
‘Sara slapped Omid.’</p> | <p>c. <i>Omid az Sārā sili xord</i>
Omid from Sara slap stroke
‘Omid was slapped by Sara.’</p> |
|-----|---|--|

On the basis of comparable data, several studies have attempted to outline a compositional analysis of Persian CPs. Despite substantial variations in their accounts, they all rely on the assumption that the contribution of the verb and the

² Abbreviations: CLP: Clitic pronoun, DOM: Differential Object Marker, EZ: Ezāfe.

non-verbal element is defined *a priori* and remains consistent through all their combinations to form a CP. Folli *et al.* (2005) and Megerdooian (2002, 2012) are recent examples of such approaches. Adopting Hale and Keyser (2002)'s "constructionalist" theory of argument structure, the authors claim that the syntactic and the semantic properties of the CP are derived from the abstract syntactic structure in which the components of the CP are inserted.

For Folli *et al.* (2005) and Megerdooian (2002), the verb realizes the verbal head *v* and determines whether an external argument (i.e. Agent) is projected³, regardless of the properties of the non-verbal element⁴. This explains the differences in argument mapping between (1b) and (1c) above. The verb furthermore determines the event type and the aspectual properties of the CP, i.e. durative vs. non-durative and eventive vs. stative: "(...) normally the eventiveness of a complex predicate depends on the light verb involved and not on the non-verbal element (Folli *et al.* 2005, p. 1379)". This in turn explains the contrast between *be yād dāštan* (Lit. 'to have in one's memory') and *be yād āvardan* (Lit. 'to bring to one's memory'): Both CPs mean 'to remember', however the first one has a stative reading while the second one denotes an event.

The non-verbal element, on the other hand, determines the Aktionsart (telic vs. atelic)⁵ and the referential properties (i.e. the lexical meaning) of the CP. CPs formed with adjectives, PPs, particles and eventive nouns are telic, while those formed with a non-eventive noun are atelic (Folli *et al.* 2005, p. 1386).

2. Problems Faced by Compositional Accounts

Whatever 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 CP components is consistent through all their combinations and can be defined *a priori*. However, as will be shown in this section, the same verb can give rise to different types of CPs with respect to their agentivity and eventive properties. Likewise, the non-verbal element's contribution can vary through its combinations with different verbs.

2.1. The Non-consistent Contribution of the CP Components

Contrary to what has been claimed in the above-mentioned studies, a given verb can form both agentive and non-agentive, as well as stative and eventive predicates. For instance, the verb *zadan* 'to hit' is generally considered as agentive and eventive. However, it can also give rise to "unaccusative" (or passive-like) CPs, like *yax zadan* 'to freeze' (Lit. 'ice hit') or *zang zadan* 'to go rusty' (Lit. 'rust hit'). The

³ "(...) the Agent-selecting properties of any given light verb are consistent across all Complex Predicates formed with a given LV." (Folli *et al.* 2005, p. 1376)

⁴ "(...) it is clear that agentivity is a property of the LV in the CP, and never depends on the nature of the NV element selected." (Folli *et al.* 2005, p. 1377)

⁵ "(...) the telicity of the CP is dependent on the non verbal element involved, in a very transparent fashion." (Folli *et al.* 2005, p. 1374)

same holds for *gereftan* ‘to take’ and *kardan* ‘to do’, which, apart from agentive CPs, e.g. *duš gereftan* ‘to take a shower’ (Lit. ‘shower take’) and *kār kardan* ‘to work’ (Lit. ‘work do’), also form “unaccusative” CPs, such as *ātaš gereftan* ‘to take fire’ (Lit. ‘fire take’), *ādat kardan* ‘to get used to’ (Lit. ‘habit do’) and *dard kardan* ‘to ache’ (Lit. ‘pain do’).

It should be noted at this point that the only criterion to which Folli *et al.* (2005) resort in order to verify the agentivity of Persian CPs is a semantic one, namely their compatibility with the adverb *amdan* ‘intentionally’: Agentive CPs are compatible with *amdan*, while non-agentive ones exclude it. The following examples show the contrast between CPs formed with *zadan* ‘to hit’ and *gereftan* ‘to take’ with respect to *amdan* and thus illustrate the fact that the same verb can give rise to both agentive and non-agentive CPs.

- (2) a. *Maryam amdan harf zad*
Maryam intentionally speech hit
‘Maryam talked intentionally.’
- b. **Āb amdan yax zad*
water intentionally ice hit
(Lit.) ‘The water froze intentionally.’
- (3) a. *Maryam amdan az Omid aks gereft*
Maryam intentionally from Omid picture took
‘Maryam took pictures of Omid intentionally.’
- b. **Dast=am amdan dard gereft*
hand=CLP.1S intentionally ache took
(Lit.) ‘My hand ached intentionally.’

However, Hale and Keyser (2002) define agentivity on syntactic grounds only. The Agent is the argument occupying the position of the Specifier of *v*, i.e. the external argument. Under this assumption, typical agentive properties such as volitionality and animacy are merely canonical interpretative properties associated to a syntactic position and consequently, all external arguments do not necessarily display these properties. Given the fact that Folli *et al.* (2005)’s criterion only singles out volitional external arguments, additional syntactic criteria are required. As shown by Samvelian (2006), the most reliable criterion to identify non-agentive CPs is their incompatibility with *=rā*, the definite/specific DO-marker in Persian⁶. The nominal element in “agentive” CPs can be *rā*-marked under certain circumstances, but never in non-agentive CPs. The examples (4) and (5) illustrate the contrast between *harf zadan* ‘to talk’ and *aks gereftan* ‘to take pictures’, agentive CPs, on one hand and *yax zadan* ‘to freeze’ and *dard gereftan* ‘to ache’, non-agentive CPs, on the other hand with respect to *rā*-marking. Thus, regardless of the criterion one resorts to, *yax zadan* and *dard gereftan* are univocally identified as non-agentive CPs.

⁶ For details on Differentiel Object Marking in Persian, which is realized by the enclitic *=rā*, see Lazard (1982) et Meunier and Samvelian (1997).

This shows the impossibility to determine *a priori* what kind of predicates a verb forms with respect to agentivity.

- (4) a. *Maryam in harf=rā zad*
Maryam this talk=DOM hit
'Maryam told this.'
- b. **āb in yax=rā zad*
water this ice=DOM hit
- (5) a. *Maryam in aks=rā az Omid gereft*
Maryam this picture=DOM from Omid took
'Maryam took this picture of Omid.'
- b. **Dast=am in dard=rā gereft*
hand=CLP.1S this ache=DOM took

The verbal contribution is not consistent either with respect to the eventive properties of the CP. Again, the same verb can give rise to both stative and eventive (dynamic) CPs. For instance, contrary to what is claimed by Folli *et al.* (2005, p. 1378), the verb *dāštan* 'to have' is not invariably stative and can produce eventive (dynamic) predicates such as *ersāl dāštan* 'to send' (Lit. 'sending have'), *taqdim dāštan* 'to offer' (Lit. 'offering have') and *e'lām dāštan* 'to announce' (Lit. 'announcing have')⁷. Table 0.1 illustrates the diversity of CPs formed with the same verb with respect to both agentive and eventive properties. Each line corresponds to a different verb.

As mentioned previously, Folli *et al.* (2005) consider the non-verbal element to be responsible for the telicity of the CP. Adjectives, PPs, particles and eventive nouns are claimed to form telic CPs, while non-eventive nouns give rise to atelic CPs. However, the contribution of the non-verbal element also turns out to be inconsistent. For instance, adjectives and PPs can as well form atelic CPs, e.g. *lāzem dāštan* 'to need' (Lit. 'necessary have'), *penhān dāštan* 'to keep hidden' (Lit. 'hidden have'), *be masxare gereftan* 'to make fun of' (Lit. 'to mockery take'). Inversely, non-eventive nouns can give rise to telic CPs, *pust andāxtan* 'to slough off' (Lit. 'skin throw').

- (6) a. *Maryam sāl-hā pul lāzem dāšt*
Maryam year-PL money necessary had
'Maryam needed money for years.'
- b. #*Maryam dar panj daqiqe pul lāzem dāšt*
Maryam in five minute money necessary had
(Lit.) 'Maryam needed money in five minutes.'

⁷ 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 CPs, see Samvelian (2012, pp. 114–130).

Subject agentivity		Event type	
Agentive	Non-agentive	Processive	Stative
<i>varaḡ zadan</i> 'to browse'	<i>kapak zadan</i> 'to go mouldy'	<i>rang zadan</i> 'to paint'	<i>barq zadan</i> 'to sparkle'
<i>farmān dādan</i> 'to order'	<i>bu dādan</i> 'to smell'	<i>anjām dādan</i> 'to accomplish'	<i>ma'ni dādan</i> 'to mean'
<i>jāru kardan</i> 'to broom'	<i>rošd kardan</i> 'to grow'	<i>ta'mir kardan</i> 'to repair'	<i>dard kardan</i> 'to ache'
<i>tasmim gereftan</i> 'to decide'	<i>anjām gereftan</i> 'to be accomplished'	<i>aks gereftan</i> 'to take a photo'	—
<i>qasam xordan</i> 'to swear'	<i>kotak xordan</i> 'to be beaten'	<i>xanjar xordan</i> 'to be stabbed'	<i>be dard xordan</i> 'to be useful'

Table 0.1: Diversity of CPs formed with the same verb

- (7) a. *Maryam sāl-hā Omid=rā be masxare gereft*
Maryam year-PL Omid=DOM to funny took
'Maryam made fun of Omid for years.'
- b. # *Maryam dar panj daqiḡe Omid=rā be masxare gereft*
Maryam in five minute Omid=DOM to funny took
(Lit.) 'Maryam made fun of Omid in five minutes.'
- (8) a. # *Mār do ruz pust andāxt*
snake two day skin threw
(Lit.) 'The snake sloughed off for two days.'
- b. *Mār dar do ruz pust andāxt*
snake in two day skin threw
'The snake sloughed off in two days.'

To conclude, *none* of the empirical generalizations stated by Folli *et al.* (2005) hold when a larger range of data is taken into account.

2.2. The Non-Predictable Semantic Content of the CP

The non-predictability of the meaning of the CP is another significant impediment to fully compositional approaches. In order for the latter to work, the meaning of the CP 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, *inter alia*), numerous Persian CPs are semantically opaque. Moreover, as shown by Samvelian (2012) and Bonami and Samvelian (2010), even in cases where a CP is semantically transparent, it is barely ever the case that its meaning is fully predictable from the meaning of its component parts.

In some cases, the CP meaning is a specialization of the predictable meaning of the combination, e.g. *čâqu zadan* ‘to stab’ (Lit. ‘knife hit’), *dast dâdan* ‘to shake hands’ (Lit. ‘hand give’), *âb dâdan* ‘to water’ (Lit. ‘water give’), *šir dâdan* ‘to breastfeed’ (Lit. ‘milk give’). Although the link between the literal and the specialized meaning is perceptible in each case, the latter cannot be straightforwardly derived from the former.

In other examples, semantic drift has taken place, either by metaphor, metonymy or ellipsis. The meaning of *gand zadan* ‘to screw up’ (Lit. ‘dirt hit’), *guš kardan* ‘to listen’ (Lit. ‘ear do’) and *zanjir zadan* ‘to flagellate’ (Lit. ‘chain hit’) can be derived via metaphor, metonymy and ellipsis respectively. Even though the meaning of these CPs is recoverable by speakers in synchrony once they learn the conventional meaning associated to them, it is not predictable *a priori* and must be learned. Moreover, in numerous other cases, the initial link is no more perceivable by speakers. For instance, *ru gereftan* ‘to become cheeky’ (Lit. ‘face take’) and *dast andâxtan* ‘to mock’ (Lit. ‘hand throw’) constitute opaque sequences in synchrony.

These facts show that the meaning of Persian CPs, even the transparent ones, is conventional in many cases and consequently has to be learned, in the same way as one has to learn the meaning of the simplex verbs in English, for instance.

3. Compositionality Revisited: A Construction-Based Approach

Relying on the observations presented in section (2), we claim that Persian CPs, at least the lexicalized ones, must be stored, exactly as lexemes are.

We nevertheless argue that the need for an inventory is not contradictory with a compositional approach, provided compositionality is defined *a posteriori*, like in Idiomatically Combining Expressions, in the sense of Nunberg *et al.* (1994). This view of Persian CPs can be developed into a Construction-based approach: 1) Each CP corresponds to a Construction. 2) CPs 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 CPs⁸.

3.1. Persian CPs as Idiomatically Combining Expressions

With respect to their compositionality, Persian CPs are comparable to Idiomatically Combining Expressions, that is “idioms whose parts carry identifiable parts of their idiomatic meanings” (Nunberg *et al.* 1994, p. 496). This means that the verb and the non-verbal element of a CP can be assigned a meaning in the context of their combination. Thus, the CP is compositional, in the sense that the meaning of the CP 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. This is the line of argumentation developed by

⁸ See Samvelian (2012) for an application of this analysis to the CPs formed with *zadan* ‘to hit’. See also Müller (2010) for a partially comparable approach within the HPSG framework.

Nunberg *et al.* (1994) to support a compositional view of expressions such as *spill the beans*. Table (0.2) illustrates this point for a set of CPs formed with *zadan* ‘to hit’. Each line contains examples of CPs where the verb can be assigned a meaning comparable to that of a lexical verb in English.

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

Table 0.2: Meanings of *zadan* in the context of its CPs

Given the meaning assigned to *zadan* and the meaning of the CP 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.

3.2. Persian CPs 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 CPs 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 CPs 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 examples of these Constructions:

- | | |
|---|--|
| <p>(9) Spreading-<i>zadan</i> Construction</p> <p><i>N0</i> (be) <i>N1</i> <i>N</i> <i>V</i></p> <p>Agent Ground Figure</p> <p>‘N0 applies N on N1’ or ‘N0 covers the surface of N1 with N’</p> | <p>(10) Instrument-<i>zadan</i> Construction</p> <p><i>N0</i> (be) <i>N1</i> <i>N</i> <i>V</i></p> <p>Agent Patient Instrument</p> <p>‘N0 accomplishes the typical action for which N is used (on N1)’</p> |
| <p>(11) Forming-<i>zadan</i> Construction</p> <p><i>N0</i> <i>N</i> <i>V</i></p> <p>Location Theme</p> <p>‘N is formed on N0’</p> | <p>(12) Slandering-<i>zadan</i> Construction</p> <p><i>N0</i> <i>be</i> <i>N1</i> <i>N</i> <i>V</i></p> <p>Slanderer Slanderee Slander</p> <p>‘N0 accuses N1 of N2’</p> |

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. 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 M. Gross (1975)’s Lexicon-Grammar notation. c) The third line provides the mapping between the semantic roles and the grammatical functions. d) The last line indicates the abstract meaning associated to the Construction as a whole. Here are examples of CPs associated to each Construction:

- (13) Spreading-*zadan* Cons: *āb* – ‘to wet’, *āhār* – ‘to starch’, *kare* – ‘to butter’, *rang* – ‘to paint’, *vāks* – ‘to polish’...
- (14) Instrument-*zadan* Cons: *jāru* – ‘to broom’, *mesvāk* – ‘to brush one’s teeth’, *otu* – ‘to iron’, *šāne* – ‘to comb’...
- (15) Forming-*zadan* Cons: *javāne* – ‘to blossom’, *juš* – ‘to sprout’, *kapak* – ‘to mouldy’, *tāval* – ‘to blister’, *zang* – ‘to go rusty’...
- (16) Slandering-*zadan* Cons: *ang* –, *bohtān* –, *ettehām* –, *tohmat* – ‘to slander’...

3.3. Networks of Constructions

Constructions can be structured in networks, reflecting different relationships such as hyponymy/hyperonymy (subtypes vs. supertypes), synonymy and valency alternations.

Subclasses and Superclasses. Some semantic classes can be grouped together into a more abstract class. In this case, the Construction associated to them is the

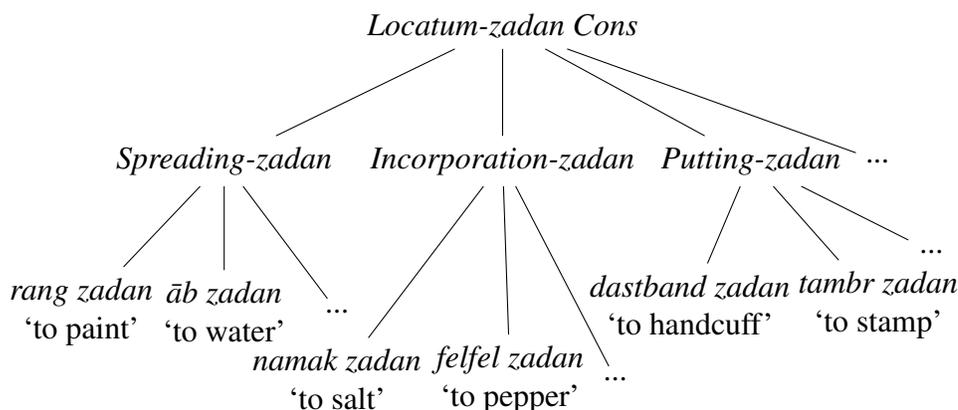


Figure 0.1: Subtypes of Locatum-zadan Construction

subtype of a less specific Construction. For instance, the Spreading-zadan Construction in (9) can be considered as a subtype of Locatum Construction. Locatum verbs (Clark and Clark 1979), e.g. *paint*, *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)’. In the case of Persian Locatum CPs, the Figure is the nominal element of the CP:

(17) **Locatum-zadan Construction**

N0 (be) *N1* *N* *zadan*
 Agent Ground Figure
 ‘N0 puts/incorporates N on/into N1’

The subtypes of the Locatum-zadan Construction, i.e. the Spreading-zadan Construction, the Incorporation-zadan Construction and the Putting-zadan Construction (cf. Table 0.1), all have an Agent and a Ground argument; the nominal element of the CP 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 predicates can generally alternate with the lexical verb *mālidan* ‘to apply’, ‘to spread’. In the CPs 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 which is an instance of the Incorporation-zadan Construction, the Figure can be confounded with the Ground or absorbed by it.

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 subtypes of a more abstract Construction, with no lexically fixed element, i.e. the Instrument Construction. Synonymy rises when the same noun occurs in the same Construction realized by different verbs.

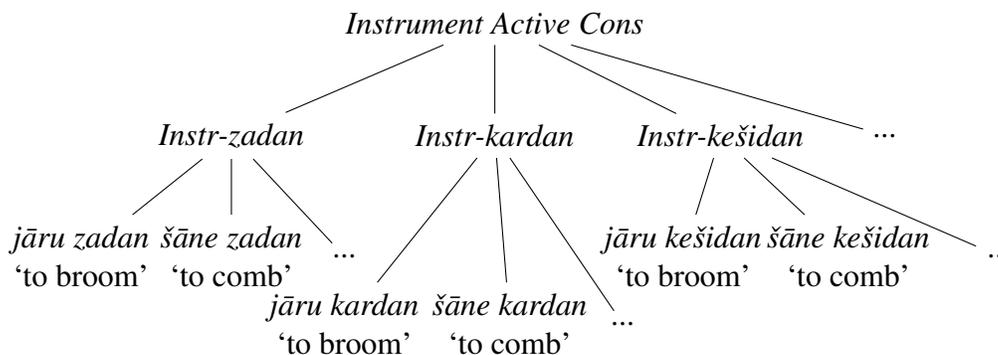


Figure 0.2: Different verbs realizing the Instrument Construction

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 CPs are often realized by the choice of the verb: *otu zadan* ‘to iron’ vs. *otu xordan* ‘to be ironed (Lit. ‘iron collide’); *âtaš zadan* ‘to set fire’, *âtaš gereftan* ‘to take fire’ (Lit. ‘fire take’).

These relations can be structured in a network using multiple inheritance hierarchies. For a partial hierarchy of Persian CPs see Figure 0.3.

4. Class Coherence, Size and Productivity

In our approach, the productivity of the Persian CPs 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 (cf. Table 0.2). 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 *es-*

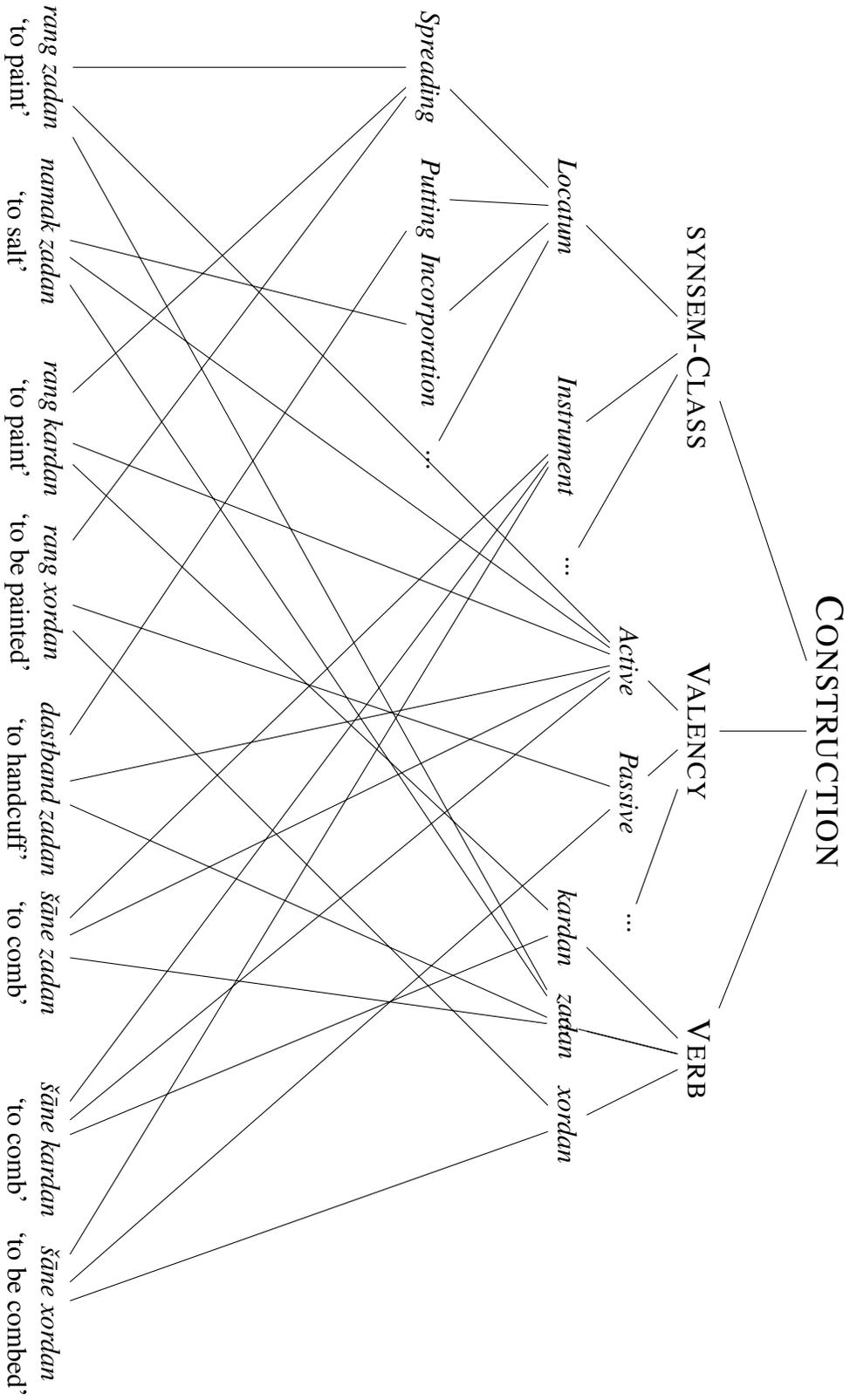


Figure 0.3: Partial hierarchy of Persian CPs

emes zadan ‘to text’, to sms’ are nevertheless created by analogical extension. We furthermore hypothesize that the productivity of Persian CPs 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 CP 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 N 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, CPs 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: Compare *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 CPs formed with *zadan*. However, some compositional classes are nevertheless 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 CPs, 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, i.e. in the case of CPs, the competing verbs to form a CP. Various verbs can be used with the same non-verbal element to realize the same “verbal concept”. This gives rise to the existence of synonymous CPs. This situation is reminiscent of construction morphology, e.g. the choice between competing affixes to form a new word.

5. Conclusion

In this paper, in line with Goldberg (1996) and Family (2006), we argued that Persian CPs 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.

Our approach goes against the majority of the previous studies on Persian complex predicates, which adopt a radically compositional perspective, built on the wrong assumption that the contribution of the CP components can be defined *a priori* and is consistent through all their combinations to form CPs. We showed that not only the lexical meaning of Persian CPs is barely ever fully predictable from the meaning of its component parts, but also that even more abstract properties, such as the argument and event structure, cannot be determined *a priori*, on the basis of solely one component of the CP regardless of the other one and the combination as a whole.

The Construction-based approach we then proposed is based on the assumption that despite their idiomacity, Persian CPs 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.

However, the issue of the productivity of Persian CPs cannot be adequately investigated without taking into account data from usage and without resorting to quantitative methods comparable to those used in morphology, which we intend to undertake in future work.

References

- Bonami, Olivier and Samvelian, Pollet, 2010. *Persian complex predicates: Lexeme formation by itself*. Paper presented at Septièmes Décembrettes Morphology Conference, Toulouse.
- Clark, Eve V. and Clark, Herbert H., 1979. *When Nouns Surface as Verbs*. *Language* 55(4):767–811.
- Family, Neiloufar, 2006. *Explorations of Semantic Space: The Case of Light Verb Constructions in Persian*. Ph.D. thesis, EHESS.
- Folli, Raffaella, Harley, Heidi, and Karimi, Simin, 2005. *Determinants of event type in Persian complex predicates*. *Lingua* 115:1365–1401.
- 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.

- Gross, Maurice, 1975. *Méthodes en syntaxe : régime des constructions complétives*. Paris: Hermann.
- Hale, K. L. and Keyser, S. J., 2002. *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: MIT Press.
- Karimi-Doostan, Gholamhossein, 1997. *Light Verb Constructions in Persian*. Ph.D. thesis, University of Essex.
- Kay, Paul and Fillmore, Charles J., 1999. *Grammatical constructions and linguistic generalizations: The What's X doing Y? construction*. *Language* 75(1–33).
- Khanlari, Parviz, 1986. *Tarix-e zabân-e farsi (A History of the Persian Language)*. Editions Nashr-e Now.
- Lazard, Gilbert, 1982. *Le morphème râ en persan et les relations actancielles*. *Bulletin de la Société de Linguistique de Paris* 77(1):177–207.
- Megerdooomian, Karine, 2002. *Beyond Words and Phrases: A Unified Theory of Predicate Composition*. Ph.D. thesis, University of Southern California.
- Megerdooomian, Karine, 2012. *The status of the nominal in Persian complex predicates*. *Natural Language and Linguistic Theory* 30(1):179–216.
- Meunier, Annie and Samvelian, Pollet, 1997. *La postposition râ en persan : son rôle dans la détermination et sa fonction discursive*. *Cahiers de grammaire* 22:187–231.
- Müller, Stefan, 2010. *Persian Complex Predicates and the Limits of Inheritance-Based Analyses*. *Journal of Linguistics* 46(2):601–655.
- Nunberg, Geoffrey, Sag, Ivan A., and Wasow, Thomas, 1994. *Idioms*. *Language* 70:491–538.
- Sadeghi, Ali Ashraf, 1993. *On denominative verbs in Persian*. In *Farsi Language and the Language of Science*, 236–246. Tehran: University Press.
- Samvelian, Pollet, 2006. *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 Faghiri, Pegah, 2013. *Introducing PersPred, a Syntactic and Semantic Database for Persian Complex Predicates*. In *Proceedings of the Workshop on Multiword Expressions*. Atlanta, Georgia, USA.

Pollet Samvelian and Pegah Faghiri
 Institut de Linguistique et de Phonétique Générale et Appliquée
 Université Sorbonne Nouvelle & CNRS
 19 rue des Bernardins, 75005 Paris (FR)

pollet.samvelian@univ-paris3.fr & pegah.faghiri@univ-paris3.fr