Persian complex predicates: Lexeme formation by itself

Olivier Bonami
Université Paris-Sorbonne & IUF
Laboratoire de Linguistique Formelle (UMR7110)
olivier.bonami@paris-sorbonne.fr

Pollet Samvelian
Université Sorbonne Nouvelle
Mondes Iranien et Indien (UMR7528)
pollet.samvelian@univ-paris3.fr

Septièmes Décembrettes, 3 décembre 2010

In Persian there is no productive morphological lexeme formation process outputting verbs. When they need to refer to a new event type, speakers resort to complex predicates (CPs). We argue that while CPs are clearly multiword expressions, CP formation has all the trappings of a lexeme formation process, and should be treated as such. Thus the class of verbal lexemes is open, but all new lexemes are multiword expressions rather than simplex words. We then propose an HPSG analysis whose key ingredient is a set of lexeme formation rules turning a noun into a verb subcategorizing for that noun.

About a dozen verbs are used in CPs, and nouns can either be predicative nouns or concrete nouns. For ease of reference we group in (1) all examples discussed in this abstract, sorted by verbs.

(1)

a.	*dast* hand	*andâxtan* throw	‘mock’	l.	čaqu knife	*zadan* hit	‘stab’
b.	*anjâm* accomplishment	*dâdan* give	‘realize’	m.	*dast* hand	*zadan* hit	‘touch, applaud’
c.	*dast* hand	*dâdan* give	‘shake hands’	n.	*faryâd* scream	*zadan* hit	‘scream’
d.	*tekân* movement	*dâdan* give	‘cause to move’	o.	*piano* piano	*zadan* hit	‘play the’ piano
e.	*dust* friend	*daštan* have	‘like’	p.	*qor* complaint	*zadan* hit	‘complain’
f.	*fekr* thought	*kardan* do	‘think’	q.	*sili* slap	*zadan* ‘hit	‘slap’
g.	*guš* ear	*kardan* do	‘listen’	r.	*šane* comb	*zadan* hit	‘comb’
h.	*dast* hand	*xordan* strike	‘be started on sth.’	s.	*šeyhe* neigh	*zadan* hit	‘neigh’
i.	juš joint	*xordan* strike	‘bind’	t.	*tašar* admonition	*zadan* hit	‘reprimand’
j.	*tekân* movement	*xordan* strike	‘be moved’	u.	*telefon* phone	*zadan* hit	‘phone’
k.	*xat* scratch	*xordan* strike	‘be scratched’	v.	*tohmat* slander	*zadan* hit	‘slander’

1 Evidence for multi-word status

The two elements in a CP are clearly separate syntactic atoms. All inflection occurs on the verb: the negative prefix occurs before the verb (2a)[1] , and the two elements can be separated by the future auxiliary (2b). Object pronominal affixes can attach to the noun (3a) in a CP, just as they can attach

[1] Abbreviations in glosses: DDO = definite direct object marker; EZ = *Ezafe* particle; NEG = negation.

generally to a complement (3b). The noun and verb can be separated by adverbs (3c). Both the nouns and verbs can be coordinated (4), and the noun can be extracted (5). Finally if the complex predicate is sufficiently compositional the noun can head a complex NP (6). These observations highlight the fact that the syntactic properties of complex predicates are identical to those of combinations of a verb with an object NP. While there is a tendency for the noun in a CP to be more cohesive with the verb than a bare direct object is (in terms of word order, stress, differential object marking, pronominal affix placement), there is no categorical syntactic contrast between the two types of sequences (*pace* Karimi-Doostan, 1997; Goldberg, 2003).

(2) a. Maryam Omid=râ dust na-dâr-ad
Maryam Omid=DDO friend NEG-have-3S
'Maryam does not like Omid.'

b. Maryam Omid=râ dust xâh-ad dâšt
Maryam Omid=DDO friend want-3S had
'Maryam will like Omid.'

(3) a. Dust=aš dâr-am
friend=3S have-1S
'I like her/him/it.'

b. Be bâzâr=aš bord.
to bazar=3S took
'(S)he took it to the bazar.'

c. Maryam Omid=râ dust aslan na-dâr-ad
Maryam Omid=DDO friend absolutely NEG-have-3S
'Maryam does not like Omid at all.'

(4) a. Maryam mu-hâ=yaš=râ bros va šâne zad
Maryam hair-PL=3S=DDO brush and comb hit
'Maryam brushed and combed her hair.'

b. Omid sili zad va xord.
Omid slap hit and strike
'Omid gave and received slaps.'

(5) Dust to Maryam=râ dâr-i?
friend you Maryam=DDO have-2S
'Is that Maryam whom you like?'

(6) Maryam [xabar=e marg=e Omid]=râ be mâ dâd.
Maryam news=EZ death=EZ Omid=DDO to us gave
'Maryam told us about Omid's death. (litt. gave us the news of Omid's death)'

2 Evidence for lexemic status

While complex predicates are multi-word combinations, the combination as a whole should be seen as the exponent of a single lexeme. Such an analysis is evidently needed in cases where the meaning of the complex predicate is opaque. The new idea we want to defend here is that productive complex predicate formation is a case of lexeme formation. We provide four arguments to this effect.

CPs are lexicalized

N-V combinations are subject to various levels of lexicalization, in a way that closely parallels what is seen with lexemes formed by morphological means. It is barely ever the case that the meaning of a CP is fully predictable from the meaning of its component parts—(1d, 1o) are good but rather isolated candidates. In many cases the CP meaning is a specialization of the predictable meaning of the combination (1c, 1l, 1m, 1r), but this particular specialization has to be learned. In other examples

semantic drift has taken place; the link between the compositional meaning and the lexicalized meaning is sometimes still recoverable synchronically (1g, 1h, 1u), sometimes not (1a, 1e). Analogy often plays an important role in motivating new lexicalizations: in (1n, 1p, 1s, 1t, 1v) the CP is formed by analogy with preexisting combinations such as (1l, 1q), not by specialization or drift from a nonlexicalized combination. Finally, even when the contribution of the verb to the CP meaning is clear, there is quite often no semantic justification for the choice of a particular verb (1b, 1f)—a situation familiar from support verb constructions, but also from affix rivalry situations.

CPs feed lexeme formation rules

N-V combinations serve as inputs to further lexeme formation rules. We give two examples of a very widespread phenomenon. (i) the suffix *-i* forms abilitative adjectives from verbs, e.g. *xordan* 'eat' > *xordani* 'edible' (and by further conversion > *xordani* 'food'). This suffix is found in combination with CPs, independently of whether they are compositional or not (7). (ii) perfect participles can regularly be converted to adjectives, and this process readily applies to CPs.

(7) a. dust daštan > dustdaštani
'love (1e)' 'lovely'

b. xat xordan > xatxordani
'be scratched (1k)' 'scratchable'

c. juš xordan > jušxordani
'bind (1i)' 'linkable'

(8) a. dast xordan > dastxorde
'be started on sth. (1h)' 'sullied'

b. xat xordan > xatxorde
'be scratched (1k)' 'scratched'

c. juš xordan > jušxordane
'bind' (1i) 'bound'

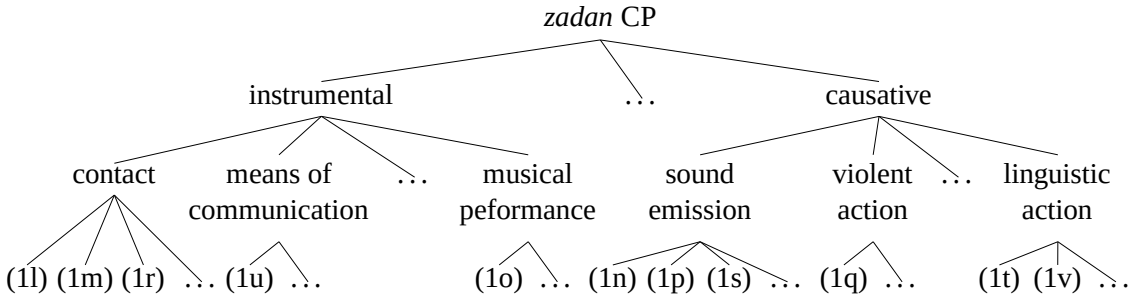
Paradigmatic groupings of CP forming verbs

Verbs used in CPs group in families with similar, if not undistinguishable, effects. *Dâdan*, *kardan* and *zadan* form instrumental or causative CPs, while *xordan*, *šodan* and *yâtan* form unaccusatives (compare 1d to 1j). Two verbs of the same family usually do not give rise to concurrent CPs, unless one of the combinations has been specialized or demotivated (compare 1c to 1m). This type of pattern closely parallels (partial) blocking effects in morphological lexeme formation (e.g. Aronoff, 1976).

Clustering of CPs based on the same verb

Complex predicates sharing the same verbal element group into clusters of related combinations with varied levels of internal coherence and of productivity, as the partial classification in (9) illustrates. However there is only a family resemblance among the clusters. Once again this is strikingly familiar to what is observed for morphological lexeme formation (e.g. Riehemann, 1998).

(9)



3 Analysis

Persian CPs are lexemes, but lexemes whose exponents take the form of combinations of two words. To borrow Gaeta and Ricca (2009)'s vocabulary, they are [+lexical,−morphological] constructions. Most existing analyses of Persian CPs are problematic because they confuse the two dimensions of analysis, and argue that PCs are words (e.g. Karimi-Doostan, 1997), phrases (e.g. Ghomeshi and Massam, 1994; Folli et al., 2005), or 'words by default' (Goldberg, 2003).

For the representation of individual PC lexemes we adapt the analysis set forth by (Müller, in press) (10): the PC enters syntax as a word form of a verbal lexeme which subcategorizes for a specific noun through the dedicated Lexical IDentifier (LID) feature (Sag, 2007; Spencer, 2004), which we

redefine as individuating lexemes via their morphological paradigm type (MP) and main semantic relation (MREL). The actual combination of verb and noun is then a matter of regular syntax.

(10) Lexical entry for the lexeme *dust dâštan* 'to like/love'.

$$\begin{bmatrix} \text{LID} & \begin{bmatrix} \text{MP} & \begin{bmatrix} \textit{dâštan} & \\ \text{STEM1} & \text{dar} \\ \text{STEM2} & \text{daʃt} \end{bmatrix} \\ \text{MREL} & \textbf{love}(i,j) \end{bmatrix} \\ \text{ARG-ST} & \left\langle \text{NP}_i, \text{NP}_j, \text{N}\begin{bmatrix} \text{LEX} & + \\ \text{LID}\,|\,\text{MP} & \textit{dust} \end{bmatrix} \right\rangle \end{bmatrix}$$

New CPs are the product of lexeme formation rules (LFRs) such as the one in (11) for intransitive instrumental CPs using *zadan*. Just like a morphological LFR, (11) turns a noun into a verb, which denotes an event type involving the use by an agent of an instance of this noun as an instrument. Two features of the new verb are unusual. First, it shares its morphological paradigm with the lexeme *zadan*, although it has a different semantics and thus a different LID. Second, it selects as a complement for a word with the same LID as the input word—thus in effect, the LFR turns a noun into a verb selecting for that noun.

(11)
$$\begin{bmatrix} \textit{lexeme} & \\ \text{CAT} & \begin{bmatrix} \textit{noun} & \\ \text{LID} & \boxed{1} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & j \\ \text{RELS} & \mathbf{P}(j) \end{bmatrix} \end{bmatrix} \mapsto \begin{bmatrix} \textit{lexeme} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{LID} & \begin{bmatrix} \text{MP} & \textit{zadan} \\ \text{MREL} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & e \\ \text{RELS} & \exists j[\mathbf{P}(j) \wedge \boxed{2}\mathbf{use}(e,i,j)] \end{bmatrix} \\ \text{ARG-ST} & \left\langle \text{NP}_i, \begin{bmatrix} \text{LEX} & + \\ \text{CAT} & \begin{bmatrix} \textit{noun} & \\ \text{LID} & \boxed{1} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Since this is a standard HPSG lexeme-to-lexeme rule, familiar analytic techniques can be applied to account for the rest of the properties. LFRs are organized in an multiple inheritance hierarchy (Riehemann, 1998), where information shared by rules based on the same verb, or having the same semantic effect, can be factored out as properties of a common supertype. Lexicalized CPs are elements of the lexical hierarchy with a frozen, non-compositional semantics (Koenig, 1999). Finally this analysis integrates readily with existing analyses of Persian morphology and syntax in HPSG.

References

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge: MIT Press.

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

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

Ghomeshi, J. and Massam, D. (1994). 'Lexical/syntactic relations without projection'. *Linguistic Analysis*, 23:175–217.

Goldberg, A. (2003). 'Words by default: Inheritance and the Persian complex predicate'. In E. J. Francis and L. A. Michaelis (eds.), *Mismatch : Form–function incongruity and the architecture of grammar*. Stanford: CSLI Publications, 117–146.

Karimi-Doostan, G. (1997). *Light verb constructions in Persian and Kurdish*. Ph.D. thesis, University of Essex.

Koenig, J.-P. (1999). *Lexical relations*. Stanford: CSLI Publications.

Müller, S. (in press). 'Persian complex predicates and the limits of inheritance-based analyses'. *Journal of Linguistics*, 46.

Riehemann, S. (1998). 'Type-based derivational morphology'. *Journal of Comparative Germanic Linguistics*, 2:49–77.

Sag, I. A. (2007). 'Sign-based construction grammar: An informal synopsis'. Ms., Stanford University.

Spencer, A. (2004). 'Generalized paradigm function morphology'. Ms., University of Essex.