The diversity of inflectional periphrasis in Persian[1]

OLIVIER BONAMI

Université Paris-Sorbonne, Institut Universitaire de France, Laboratoire de Linguistique Formelle (CNRS & U. Paris Diderot)

POLLET SAMVELIAN

Université Sorbonne Nouvelle, Mondes Iranien et Indien (CNRS, Inalco, EPHE & U. Sorbonne Nouvelle)

Modern Persian conjugation makes use of five periphrastic constructions with typologically divergent properties. This makes the Persian conjugation system an ideal testing ground for theories of inflectional periphrasis, since different types of periphrasis can be compared within the frame of a single grammatical system.

We present contrasting analyses of the five constructions within the general framework of a lexicalist constraint-based grammatical architecture (Pollard & Sag, 1994) embedding an inferential and realizational view of inflectional morphology (Stump, 2001). We argue that the perfect periphrase can only be accounted for assuming that the periphrase literally fills a cell in the inflectional paradigm, and provide a formal account relying on using valence for exponence. On the other hand, other periphrastic constructions are best handled using standard tools of either morphology or syntax. The overall conclusion is that not all constructions that qualify as periphrastic inflection from the point of view of typology should receive the same type of analysis in an explicit formal grammar.

1. INTRODUCTION

Two quite different definitions can be given for the notion of *inflectional periphrasis*. On a more permissive definition, a syntactic construction is a case of inflectional periphrasis if it serves as the realization of a property that is typically thought to be inflectional (see e.g. Spencer 2006, Brown et al. 2012). Such a definition is often implicit in descriptive grammars, and is definitely useful for typology, allowing one to see periphrasis as a gradient property, going from open syntactic combination to full morphologization. We will henceforth call this the typological definition of periphrasis.

A narrower definition takes periphrasis to be a type of analysis for a given construction. In this sense, a construction is seen as periphrastic if it is a multi-word construction that interacts with inflectional morphology in such a way that it is best integrated in the inflectional paradigm (see e.g. Hockett 1958, Haspelmath 2000, Ackerman & Stump 2004). Under this view, an inflectional periphrase is

the realization of a morphosyntactic feature bundle on a lexeme, just in the same sense as an inflected word is. Such a stricter definition is directly relevant to lexicalist models of morphosyntax, which need to either be adapted to encompass such constructions (see e.g. Börjars et al. 1997, Ackerman & Webelhuth 1998, Sadler & Spencer 2001, Ackerman & Stump 2004, Bonami & Webelhuth 2013, Blevins to appear) or to argue against their existence (see e.g. the analysis of Latin periphrastic perfects in Kiparsky 2005). We will henceforth call this the formal definition of periphrasis.

Deciding which periphrastic constructions in the typological sense need to receive periphrastic analyses in the formal sense can be difficult. Ackerman & Stump (2004) and Haspelmath (2000) propose general empirical criteria, but the relevance of such criteria depends on assumptions on morphology, syntax, and the morphology-syntax interface that are hardly independent of the analysis of the phenomena at hand, as we will see in section 7. Our goal in this paper is to disentangle some of the issues by looking at an inflectional system involving a diverse set of periphrastic constructions in the typological sense. We will try to determine which of those can be reduced to normal syntax or normal morphology, and provide an explicit, periphrastic in the formal sense, analysis for the remaining cases. Our general strategy is therefore conservative: we characterize a construction as formally periphrastic only if there is no satisfactory purely morphological or purely syntactic analysis. Comparing our analysis of the formally periphrastic constructions to that of the other cases will shed light on the more general typology of inflectional periphrasis. The analysis will rely on the association of a specific syntactic framework (Head-driven Phrase Structure Grammar; Pollard & Sag (1994)) and a specific framework for inflectional morphology (Paradigm Function Morphology; Stump (2001)). This association is particularly promising for the problem at hand, for two main reasons: first, HPSG's use of a uniform, rich ontology for all aspects of linguistic description makes it easier to define a tight interface with dense information flow between the syntactic and morphological components, as necessitated by the analysis of periphrases. Second, although the two frameworks rely on different ontologies, PFM is the only extant formally explicit theory of morphology that can readily be recast in an ontology of typed feature structure descriptions.[2]

Grammars of Persian (e.g. Lazard et al. 2006) distinguish four conjugational periphrastic construction types. The passive construction is based on an inflected form of *šodan* 'become' preceded by a perfect participle (1). So-called 'perfect' forms are based on an inflected form of *budan* 'be' preceded by a perfect participle (2)—the auxiliary is a full word (2a), or a clitic (2b) depending on tense and mood. The future is formed with a special present tense form of *xâstan* 'want' followed by a short infinitive (3). Finally, the progressive is based on an inflected form of *dâštan* 'have' followed by a finite form (4).[3]

2

(1) In   tâblo   foruxte mi-šav-ad.   (Passive)
     this painting sell.PFP UNBD-become.PRS-3SG
     'This painting is sold.'

(2) a. Maryam in   tâblo=râ   foruxte bud.   (Perfect)
        Maryam this painting=DDO sell.PFP COP.PST.3SG
        'Maryam had sold this painting.'

     b. Maryam in   tâblo=râ   foruxte=ast.
        Maryam this painting=DDO sell.PFP=COP.PRS.3SG
        'Maryam has sold this painting.'

(3) Maryam in tâblo=râ   xâh-ad   foruxt.   (Future)
     Maryam this painting=DDO want.PRS-3SG sell[SINF]
     'Maryam will sell this painting'

(4) Maryam dâr-ad   in tâblo=râ   mi-foruš-ad.
     Maryam have.PRS-3SG this painting=DDO UNBD-sell.PRS-3SG
     'Maryam is selling this painting.'   (Progressive)

The differing properties of these four types of periphrasis result from different origins as finite, infinitival or participial complements, and different degrees of grammaticalization, going from the quasi-analytic passive to the recently morphologized present perfect.

The periphrastic status of a construction can only be evaluated by comparison with characteristic open syntactic constructions and morphological combinations in the relevant language. We thus start in section 2 with an overview of Persian clausal syntax and conjugation, including sketches of formal analyses. Section 3 deals with the passive construction, section 4 with the 'perfect', section 5 with the future, and section 6 with the progressive. Section 7 discusses the position of the Persian constructions in the typology of periphrastic constructions.

Since periphrastic constructions occupy an uncertain and still ill-understood position on the border between typical syntax and typical morphology, the basic descriptive vocabulary tends to be confusing. Part of the problem stems from the fact that two basic intuitions come into conflict when we think of periphrasis: the intuition that inflection is a subcomponent of morphology, and the intuition that morphology deals with words defined as syntactic atoms. Some authors are happy to abandon the latter intuition, and take periphrasis to be part of morphology (e.g. Ackerman et al. 2011: 332) or at the intersection of morphology and syntax (e.g. Brown et al. 2012: 247). Although this is arguably a matter of taste, we find this use of the terminology confusing. In this paper we take the opposite decision. We define inflection as the component of grammar responsible for the realization of lexemes in context. That realization may use morphological means and take the form of a word (we call this synthetic inflection), or it may use a combination of morphological and syntactic means and take the form of a combination of multiple words; we call this periphrastic inflection, and we call the combination of words a periphrase. Notice that under this definition, the term 'inflectional morphology'

3

is not synonymous with 'inflection', but rather denotes the use of morphology for the expression of inflection.[4]

2. RELEVANT BASIC FEATURES OF PERSIAN GRAMMAR

2.1 *Grammatical functions*

Alignment is clearly accusative in Persian. Subjects and objects contrast most notably on three features. First, subjects, but not objects, trigger person and number agreement on finite verbs. Second, objects, but not subjects, may be realized as enclitic pronouns (5).[5] Third, objects, but not subjects, may carry the clitic =*râ*. The distribution of =*râ* is quite complex, and it may appear on nonobject constituents, but it never attaches to a subject (Karimi 1996, Ghomeshi 1997, Samvelian 2006-2007). When attached to an object, it marks it as definite/specific (6).

(5)   a. Omid in tablô=râ xarid.
        Omid this painting=DDO buy.PST.3SG
        'Omid bought this painting.'

       b. Omid xarid=aš.
        Omid buy.PST.3SG=3SG
        'Omid bought it.'

(6)   a. Omid tablô=râ xarid.
        Omid painting=DDO buy.PST.3SG
        'Omid bought the painting.'

       b. Omid tablô xarid.
        Omid painting buy.PST.3SG
        'Omid bought a painting/paintings.'

Persian is extensively *pro*-drop, and pronominal subjects are seldom realized overtly, except for contrastive, emphatic or reflexive purposes (7).

(7)   a. Tablô=râ xarid-am.
        painting=DDO buy.PST-1SG
        'I bought the painting.'

       b. Man tablô=râ xarid-am (na Maryam).
        I painting=DDO buy.PST-1.SG not Maryam
        'I bought the painting (not Maryam).'

       c. Xod-am tablô=râ xarid-am (na Maryam).
        self-CL.1SG painting=DDO buy.PST-1SG not Maryam
        'I myself bought the painting.'

2.2 *Basic word order*

Word order is quite free in Persian. In the written standard, dependents of the verb tend to precede it (the least marked order is SOV), except for subordinate clauses which always occur after the verb. However in colloquial Persian subjects (8a), objects (8b), all complements (8c) and most adjuncts (8d) may also occur after the verb.

(8) a. Tablô=râ   xarid   Maryam.
         painting=DDO buy.PST.3SG Maryam
         'Maryam bought the painting.'

    b. Maryam xarid   tablô=râ.
        Maryam buy.PST.3SG painting=DDO.
        'Maryam bought the painting.'

    c. Maryam tablô=râ   xarid   az   Omid.
        Maryam painting=DDO buy.PST.3SG from Omid
        'Maryam bought the painting from Omid.'

    d. Maryam in tablô=râ   xarid   sâl=e   piš.
        Maryam this painting=DDO buy.PST.3SG year=EZ before
        'Maryam bought this painting last year.'

2.3 *Subordinate clauses*

Finite complement clauses can be marked with the complementizer *ke*, but the complementizer is very easily dropped (9a). In addition, embedded valents are easily scrambled to the matrix clause (9b-d), independently of the presence or absence of the complementizer.

(9) a. Maryam mi-dân-ad   [(ke) Omid in ketâb=râ   be Sârâ
         Maryam UNBD-know.PRS-3SG   that Omid this book=DDO to Sara
         dâd].
         give.PST.3SG
         'Maryam knows that Omid gave this book to Sara.'

    b. Maryam in ketâb=râ mi-dân-ad [(ke) Omid be Sârâ dâd].

    c. Maryam be Sârâ mi-dân-ad [(ke) Omid in ketâb-râ dâd]

    d. Maryam in ketâb-râ be Sârâ mi-dân-ad [(ke) Omid dâd]

Complement clauses are almost always finite in colloquial Persian. Nonfinite complements have a restricted distribution; with a handful of exceptions, the natural translation of an English equi verb is a verb or complex predicate taking a finite complement clause with no overt subject. Optional control verbs have the same construction, whether the embedded object is coreferential with an argument of the embedding verb or not (10).

(10) a. Maryam mi-xâh-ad   [(ke) bâ Omid har ruz be sinemâ
Maryam UNBD-want.PRS-3SG (that) with Omid every day to theatre
be-rav-ad].
SBJV-go.PRS-3SG
'Maryam wants to go to theatre with Omid everyday.'

      b. Maryam mi-xâh-ad   [(ke) bâ Omid har ruz be sinemâ
Maryam UNBD-want.PRS-3SG (that) with Omid every day to theatre
be-rav-am].
SBJV-go.PRS-1SG
'Maryam wants me to go to theatre with Omid everyday.'

Predicates corresponding to raising verbs in English may have two contrasting constructions. The more common option is to have an impersonal construction with no subject (11); this is found e.g. with the verb *bâyestan* 'must', *šodan* 'become' in its modal use, the adjective *momken* 'possible' in its predicative use, or the complex predicate *be nazar âmadan* 'seem'.[6] Note that while the subject of the embedded clause may be realized before the matrix verb (12a), this is clearly a long-distance dependency rather than a raised subject, as the agreement is with the embedded predicate only (12b).

(11) Mi-šav-ad   [(ke) bačče-hâ madrase be-rav-and].
UNBD-become-3SG that child-PL school SBJV-go.PRS-3PL
'The kids may go to school.'

(12) a. Bačče-hâ mi-šav-ad   [(ke) madrase be-rav-and].
child-PL UNBD-become-3SG that school SBJV-go.PRS-3PL
'The kids may go to school.'

      b. * Bačče-hâ mi-šav-and   [(ke) madrase be-rav-and].
child-PL UNBD-become-3PL that school SBJV-go.PRS-3PL

A less common option is to use a raising construction. To our knowledge this is only found with *tavânestan* 'can'. First, the NP *bačče-hâ* in (13a) is a subject of the matrix verb since it agrees in number with it (contrast (13b)) and can not be realized within the finite complement clause (contrast (13c)). Second, and *pace* Karimi (2008), *tavânestan* is not a control verb: the subject can be inanimate and non-causal (14), in contrast to English *be able to*; and *tavânestan* can take an impersonal construction as its complement (15), a fact that is incompatible with its assigning a thematic role to a subject. The unescapable conclusion is that, while rare, raising constructions exist in Persian.

(13) a. Bačče-hâ mi-tavân-and   [(ke) madrase be-rav-and].
child-PL UNBD-can.PRS-3PL that school SBJV-go.PRS-3PL
'The kids can go to school.'

      b. * Bačče-hâ mi-tavân-ad   [(ke) madrase be-rav-and].
child-PL UNBD-can.PRS-3SG that school SBJV-go.PRS-3PL

c. * Mi-tavân-ad   [(ke) bačče-hâ madrase be-rav-and].
       UNBD-can.PRS-3SG   that child-PL   school   SBJV-go.PRS-3PL

(14) In   ketâb mi-tavân-ad   [barâ=ye hame mofid bâsh-ad]   .
      this book UNBD-can.PRS-3SG   for=EZ   all   useful be.SBJV-3SG
      'This book can be useful to anybody.'

(15) Mi-tavân-ad   [be nazar   bi-ây-ad   [ke in talâsh-hâ
      UNBD-can.PRS-3SG   to opinion SBJV-come.PRS-3SG   that this effort-PL
      bihude=ast]].
      vain=COP.PRS.3SG
      'It may seem that these efforts are vain.'

2.4 *Lexical complements*

Persian makes heavy use of constructions that fall under the general category of *complex predicates*, i.e., where the head verb combines with a non-head word some of whose arguments are realized in the clausal domain (Barjasteh 1983, Karimi 1997, Karimi-Doostan 1997). These constructions go from idioms (16) to free combinations such as copular constructions (19) through productive and compositional light verb constructions (17). What all these constructions have in common is that the arguments of the non-head word have ordering possibilities characteristic of a valent of the head. The contrast between the two copular constructions in (18) and (19) illustrates this. (18) is a phrasal complementation structure: the adjective *negarân* combines through the *ezafe* particle with its complement, and the two constituents form a phrase that can't be disrupted, and serves as the complement of the copula. By contrast, in (19), although *az Omid* clearly expresses an argument of the adjective *râzi*, the two constituents exhibit no syntactic cohesion; in particular *az Omid* may be realized in all positions that are legal for a dependent of the head verb (here the clitic copula).

(16) Maryam be Omid sar   zad.
      Maryam to Omid head strike.PST.3SG
      'Maryam paid a visit (lit. struck head) to Omid.'

(17) Be daftar=e   riâsat=e   jomhuri niz   imeyl zad-im.
      to   office=EZ presidency=EZ republic also email strike.PST-1PL
      'We also emailed (lit. struck email) the office of the presidency of the republic.'

(18) a.   Maryam negarân=e   Omid=ast.
         Maryam worried=EZ Omid=COP.PRS.3SG
         'Maryam is worried about Omid.'

     b. * Maryam Omid negarân=e=ast.
         Maryam Omid worried=EZ=COP.PRS.3SG

7

    c.   * Maryam Omid negarân=ast.
        Maryam Omid worried=COP.PRS.3SG

(19) a. Mayam râzi   az Omid=ast.
        Maryam satisfied of Omid=COP.PRS.3SG
        'Maryam is satisfied with Omid.'

    b. Maryam az Omid râzi=ast.

    c. Az Omid Maryam râzi=ast.

    d. Maryam râzi=ast az Omid.

2.5 *Verbal complex*

In a limited number of constructions, the head verb is tightly linked to a lexical complement. A case in point is that of particle verbs such as *bar dâštan* 'take' exemplified in (20): the particle-verb sequence can only be interrupted by an object clitic (21) or the future auxiliary. This suggests that the head verb forms a small phrase with its complement, which we will call a *verbal complex*, and that the range of elements that may belong to this phrase is drastically limited.

(20) a.   Omid ketâb=râ bar dâšt.
          Omid book=DDO PART have.PST.3SG
          'Omid took the book.'

    b.   * Omid bar   ketâb dâšt.
          Omid PART book have.PST.3SG

(21) a.   Bar dâr=eš!
          on have[IMP.2SG]=CL.3SG

    b.   Bar=eš   dar!
          PART=CL.3SG have[IMP.2SG]
          'Take it!'

2.6 *Topicalization*

In addition to co-valent reordering and long distance scrambling, Persian has a topicalization construction, allowing for the fronting of almost any constituent. Note that this is clearly a long-distance dependency (22a), and that it differs from scrambling in possibly affecting finite verbs (22b).

(22) a. Az   Omid   Maryam   fekr   mi-kon-am   [ke
          from   Omid   Maryam   thought UNBD-do.PRS-1SG   that
          mi-xâh-ad   tâblô   be-xar-ad].
          UNBD-want.PRS-3SG painting SBJV-buy.PRS-3SG
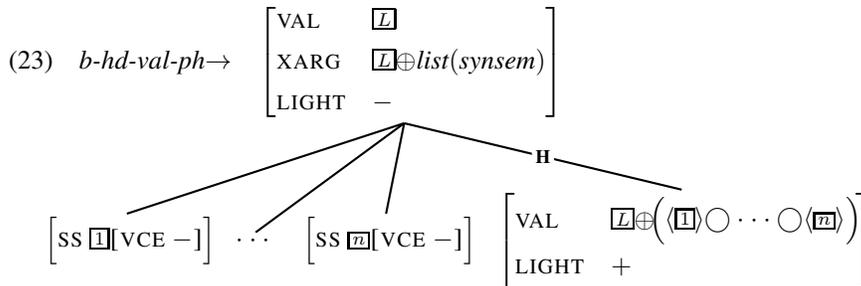   'From Omid, I think that Maryam wants to buy a painting.'

b. Dâd Maryam fekr   mi-kon-am   [in   tâblo=râ   be
   gave Maryam thought UNBD-do.PRS-1SG   this painting=DDO to
   Omid diruz].
   Omid yesterday
   'I think that Maryam gave this painting to Omid yesterday.'

2.7 *A preliminary HPSG analysis*

In this paragraph we sketch an HPSG analysis which accounts for most of the
basic features of Persian syntax discussed above. This preliminary account is
necessary to ground the analysis of periphrases in later sections, but many aspects
are open to debate, and could be altered without implying major changes in that
analysis.[7]

Word order facts and constituency tests provide no motivation for a VP/S
asymmetry in Persian: subjects and phrasal complements may be freely reordered.
The basic head-valence phrase in (23) thus realizes multiple dependents of the
head in the same local tree, as illustrated in Fig. 1, and without constraining
their relative order.[8] This is facilitated by the fact that all valents, subjects and
complements alike, are licensed by a single feature VAL, rather than separate SUBJ
and COMPS features. Wherever the grammar needs to single out subjects, this is
done through the feature XARG (for 'external argument') which distinguishes at
most one argument as externally selectable (Sag 2012). Notice that XARG is not
a valence feature (its value is not cancelled when the head combines with the
designated argument) but a head feature, as indicated in (24).

The specification in (23) leaves the possibility open that a basic head-valence
phrase be unsaturated for a subject argument. Since XARG is independently
constrained to contain at most one element (see (29) below), (23) is compatible
with two possibilities: L contains one element corresponding to the subject, and
the phrase is unsaturated; or L is empty, the phrase is saturated, and whether there
is a subject or not (i.e. whether the list concatenated with L as the value of XARG
is empty or of length one) is left to independent constraints. Unsaturated head-
valence phrases will be crucial to the analysis of raising constructions. However,
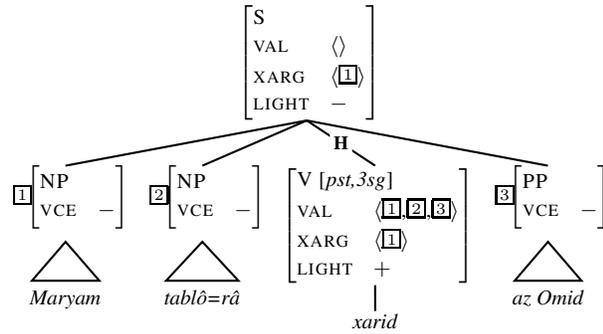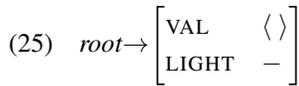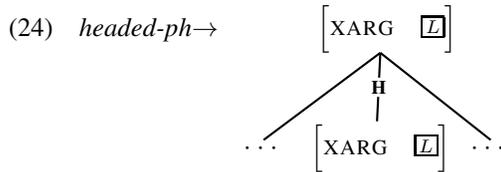as indicated in (25), root clauses are constrained to be saturated for valence.

(23) *b-hd-val-ph* →
$$\begin{bmatrix} \text{VAL} & L \\ \text{XARG} & L \oplus list(synsem) \\ \text{LIGHT} & - \end{bmatrix}$$

9

$$\begin{bmatrix} \text{S} \\ \text{VAL} \quad \langle\rangle \\ \text{XARG} \quad \langle\boxed{1}\rangle \\ \text{LIGHT} \quad - \end{bmatrix}$$

$\boxed{1}\begin{bmatrix} \text{NP} \\ \text{VCE} \quad - \end{bmatrix}$ $\boxed{2}\begin{bmatrix} \text{NP} \\ \text{VCE} \quad - \end{bmatrix}$ **H** $\begin{bmatrix} \text{V } [pst,3sg] \\ \text{VAL} \quad \langle\boxed{1},\boxed{2},\boxed{3}\rangle \\ \text{XARG} \quad \langle\boxed{1}\rangle \\ \text{LIGHT} \quad + \end{bmatrix}$ $\boxed{3}\begin{bmatrix} \text{PP} \\ \text{VCE} \quad - \end{bmatrix}$

Maryam   *tablô=râ*   *xarid*   *az Omid*

Figure 1
Analysis for (8c)

(24) *headed-ph*→ $\begin{bmatrix} \text{XARG} \quad \boxed{L} \end{bmatrix}$

H

··· $\begin{bmatrix} \text{XARG} \quad \boxed{L} \end{bmatrix}$ ···

(25) *root*→ $\begin{bmatrix} \text{VAL} \quad \langle\,\rangle \\ \text{LIGHT} \quad - \end{bmatrix}$

The binary features VCE and LIGHT are used to specify that some 'smaller' constituents have to combine closer to the head. First, words specify lexically their dependents as being either [VCE +] (verbal complex elements) or [VCE −]: in the former case they can combine with the head only inside the verbal complex, as specified in (26). Second, the feature LIGHT, adapted from Abeillé & Godard (2000), captures the syntactically relevant notion of 'lexicality'.[9] Roughly, a [LIGHT +] constituent is a word or 'small' phrase consisting of [LIGHT +] words, such as a word-level coordination; ordinary phrases are [LIGHT −]; thus the specifications for LIGHT in (23) and (26) ensure that a verbal complex can head a clause, but not the other way around, while still allowing for coordinations of verbs to head a verbal complex. The coindexation between VAL and XARG in (26) ensures that the verbal complex phrase introduces as a sister to the head the least oblique non-subject valent. The effect of (26) is illustrated in Fig. 2.
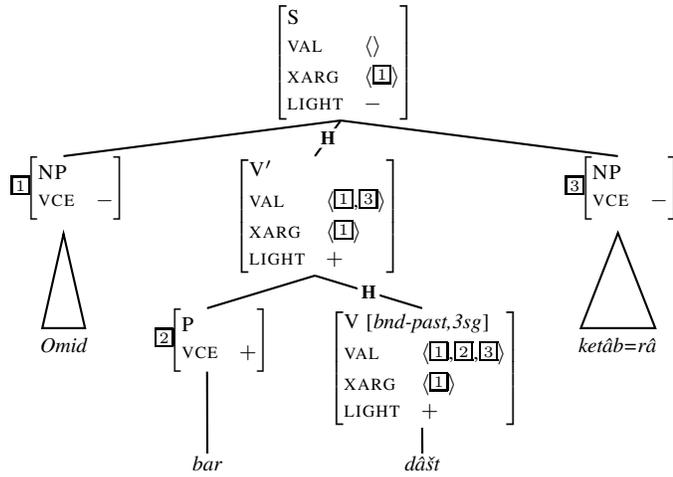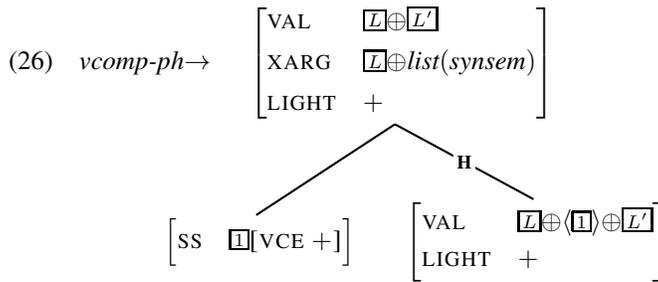
$$
\begin{bmatrix}
\text{S} \\
\text{VAL} & \langle\rangle \\
\text{XARG} & \langle\boxed{1}\rangle \\
\text{LIGHT} & -
\end{bmatrix}
$$

H′

$$
\boxed{1}\begin{bmatrix}\text{NP} \\ \text{VCE} & -\end{bmatrix}
\qquad
\begin{bmatrix}
\text{V}' \\
\text{VAL} & \langle\boxed{1},\boxed{3}\rangle \\
\text{XARG} & \langle\boxed{1}\rangle \\
\text{LIGHT} & +
\end{bmatrix}
\qquad
\boxed{3}\begin{bmatrix}\text{NP} \\ \text{VCE} & -\end{bmatrix}
$$

Omid

H

$$
\boxed{2}\begin{bmatrix}\text{P} \\ \text{VCE} & +\end{bmatrix}
\qquad
\begin{bmatrix}
\text{V}\ [\textit{bnd-past,3sg}] \\
\text{VAL} & \langle\boxed{1},\boxed{2},\boxed{3}\rangle \\
\text{XARG} & \langle\boxed{1}\rangle \\
\text{LIGHT} & +
\end{bmatrix}
$$

ketâb=râ

bar   *dâšt*

Figure 2
‘Omid took the book’

(26)   *vcomp-ph*→
$$
\begin{bmatrix}
\text{VAL} & \boxed{L}\oplus\boxed{L'} \\
\text{XARG} & \boxed{L}\oplus list(synsem) \\
\text{LIGHT} & +
\end{bmatrix}
$$

H

$$
\begin{bmatrix}\text{SS} & \boxed{1}[\text{VCE} +]\end{bmatrix}
\qquad
\begin{bmatrix}
\text{VAL} & \boxed{L}\oplus\langle\boxed{1}\rangle\oplus\boxed{L'} \\
\text{LIGHT} & +
\end{bmatrix}
$$

We do not assume that lexical ([LIGHT +]) complements of the verb are systematically part of the verbal complex, since they need not be close to the verb. The examples in (19) are licensed by a lexical entry of the copula which combines with a lexical adjective and raises its arguments, as illustrated in Fig. 3.

Our account of *pro*-drop and pronominal affixation relies on the notion of non-canonical realization of arguments put forth among others by Miller & Sag (1997). The central assumption is that whereas verbs represent in their lexical entry the list of their all their syntactic arguments (in tight but non-trivial correspondence with the list of semantic arguments of the relation they denote), only a subset of these are projected as valents to be realized as local syntactic dependents of the verb. Those arguments that do not project as valents are called *non-canonical*. This assumption is spelled out by the constraint in (27), stating that all and only canonical elements of a verb's ARG-ST list also occur on its VAL list.
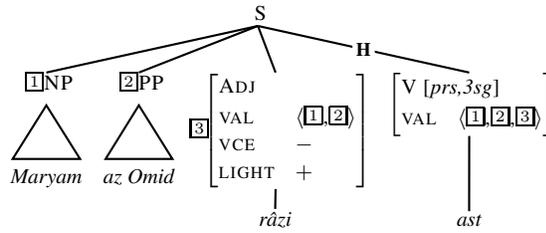
11

Figure 3
Analysis for (19b)

(27) Argument Realization

$$verb\text{-}wd \rightarrow \begin{bmatrix} \text{ARG-ST} & \boxed{L} \bigcirc list(noncanon) \\ \text{VAL} & \boxed{L} \, list(canon) \end{bmatrix}$$

Different kinds of non-canonical *synsem*s may be distinguished to account for the syntactic and semantic variability of non-phrasal (or non-local phrasal) realizations. For concreteness we distinguish: gaps (indicating locally that a dependent has been extracted), non-canonical referential pronouns (corresponding e.g. to *pro*-dropped arguments), non-canonical generic arguments (corresponding to arbitrary PRO), and non-canonical indefinites (corresponding to implicit arguments with an indefinite interpretation).

(28)
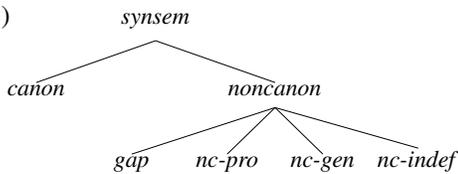
A non-canonical pronoun is a pronoun with no syntactic exponence, but which may have inflectional exponence. This happens in Persian when an object clitic is realized locally on the head verb.[10]

As Bonami & Samvelian (2008) observe, in a realizational theory of morphology which recognizes the existence of non-canonical pronouns, the possibility of *pro*-drop follows without stipulation. In a realizational theory, a feature will be realized only if there happen to be inflectional rules realizing it. Saying that pronominality is a morphosyntactic feature does not entail that it *has* to be realized by specific morphology. We thus posit that Persian has non-canonical subject pronouns, a property it shares with other *pro*-drop languages, but also with some non-*pro*-drop languages which express their subject pronouns affixally, like French (Miller 1992). Persian differs from French not in having a rule of *pro*-drop that French lacks, but in having no inflectional rule specifically expressing the pronominality of the subject; English on the other hand differs
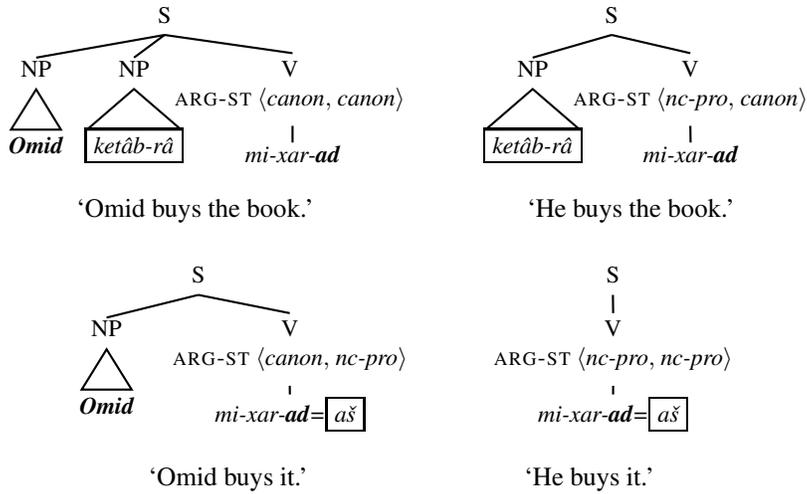
Figure 4
Canonical and non-canonical arguments

from both Persian and French in lacking non-canonical pronouns altogether, and thus requiring syntactic realization of pronominal arguments.All three languages have rules expressing subject agreement, that is, person and number properties of the subject, irrespective of whether it is canonical or non-canonical, pronominal or non-pronominal. As a result, in French, person and number of a non-canonical pronoun will be spelled out twice by inflection.

The trees in Fig. 4 illustrate the distribution of canonical and non-canonical arguments in Persian, where (syntactic or morphological) exponents of the subject are in boldface and exponents of the object are boxed.

Fig. 5 illustrates the analysis of a sentence with a null pronominal (so-called '*pro*-dropped') subject. Notice that while the verb has a syntactic subject, this subject is not realized in any way by phrasal means, and the phrase is syntactically saturated. However information on the subject is accessible at phrase level through the feature XARG.

We are now in a position to turn to the analysis of different types of verbal complementation. First, we noticed earlier that although most verbs take a subject, some, such as *bâyestan*, are subjectless. We thus assume at lexical level a basic division between *personal* and *impersonal* verbs, that is specified purely in terms of XARG: personal verbs have one element on XARG (29a), impersonal verbs have none (29b). The value of XARG has to be the initial part of the ARG-ST of verbs (30), hence the subject, if there is one, is the least oblique argument of the verb.
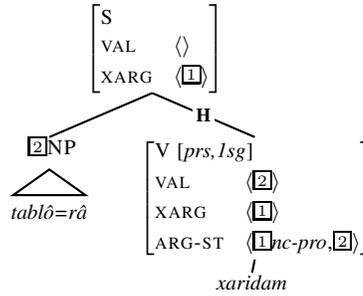
$$
\begin{bmatrix} S \\ \text{VAL} \quad \langle\rangle \\ \text{XARG} \quad \langle \boxed{1} \rangle \end{bmatrix}
$$

$$
\boxed{2}\,\text{NP} \qquad \mathbf{H}
$$

$$
\overset{\triangle}{tablô{=}râ} \qquad \begin{bmatrix} \text{V } [prs,1sg] \\ \text{VAL} \quad \langle \boxed{2} \rangle \\ \text{XARG} \quad \langle \boxed{1} \rangle \\ \text{ARG-ST} \quad \langle \boxed{1}nc\text{-}pro, \boxed{2} \rangle \end{bmatrix}
$$

$$
xaridam
$$

Figure 5
Analysis for (7a) 'I bought the painting'

(29) a. $per\text{-}vb\text{-}wd \rightarrow \begin{bmatrix} \text{XARG} \quad \langle synsem \rangle \end{bmatrix}$

   b. $imper\text{-}vb\text{-}wd \rightarrow \begin{bmatrix} \text{XARG} \quad \langle\rangle \end{bmatrix}$

(30) $vb\text{-}wd \rightarrow \begin{bmatrix} \text{XARG} \quad \boxed{L} \\ \text{ARG-ST} \quad \boxed{L} \oplus list(synsem) \end{bmatrix}$

We distinguish four subtypes of verbs taking a clause-like complement. In the simplest case, the verb selects for a finite complement saturated for its valence. This is ensured by the schematic lexical entry for *dânestan* in (31).

(31) $\begin{bmatrix} \text{LID} \qquad dânestan \\ \text{ARG-ST} \quad \left\langle \text{NP,} \begin{bmatrix} \text{HEAD} \quad \text{V}[ind] \\ \text{VAL} \quad \langle\rangle \\ \text{LIGHT} \quad - \end{bmatrix} \right\rangle \end{bmatrix}$

Although the data is compatible with other approaches, for concreteness we adopt a 'weak head' (Abeillé et al. 2006) analysis of complementizers along the lines of Tseng (2002). The complementizer *ke* takes an unmarked finite clause as its complement (32) and projects a phrase specified as [MARKING *ke*]. Moreover, *ke* shares the HEAD value of its complement and thus projects a finite clause.

(32) *ke*: $\begin{bmatrix} \text{HEAD} \quad \boxed{1} \\ \text{MARKING} \quad ke \\ \\ \text{VAL} \quad \boxed{L} \oplus \left\langle \begin{bmatrix} \text{HEAD} \quad \boxed{1}\begin{bmatrix} verb \\ \text{VFORM} \quad fin \end{bmatrix} \\ \text{MARKING} \quad none \\ \text{VAL} \quad \boxed{L} \\ \text{LIGHT} \quad - \end{bmatrix} \right\rangle \end{bmatrix}$
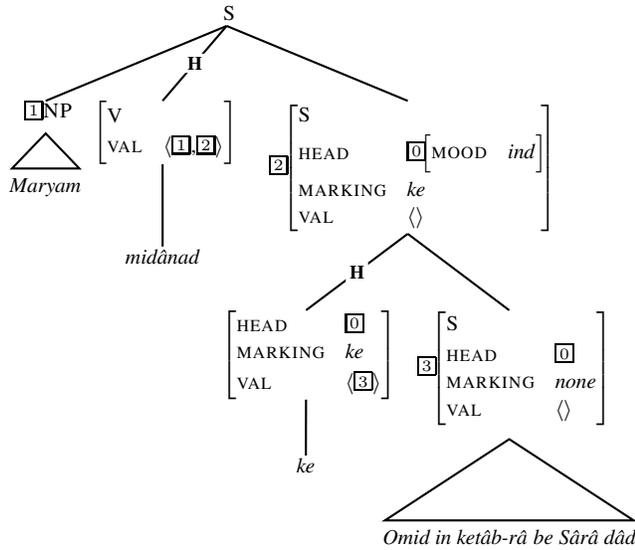
14

Figure 6
Analysis for (9a) 'Maryam knows that Omid gave this book to Sara'

The lexical entries for *dânestan* and *ke* jointly license the analysis in Fig. 6. Notice that like most verbs, *dânestan* does not constrain the MARKING value of its clausal complement, and thus will be compatible either with a *ke*-marked or an unmarked complement. Also notice that *ke* inherits the valence of its complement. Since only subjects are allowed to be unsaturated at phrase-level, this allows *ke* to mark a phrase not saturated for the subject. In the case at hand however, *dânestan* selects for a saturated complement and thus the complement of *ke* must also be saturated.

Obligatory control constructions are licensed by lexical entries very similar to (31), the only difference being that the embedded subject is constrained through XARG to be a non-canonical pronoun coindexed with a matrix argument. The lexical entry for the complex predicate *tasmim gereftan* illustrates. Since they have an empty XARG, impersonal verbs are excluded as complements of a control verb.

S

H

[1]NP
$\begin{bmatrix} V \\ \text{XARG} & \langle[1]\rangle \\ \text{VAL} & \langle[1],[2]\rangle \\ \text{ARG-ST} & \langle[1],[2]\rangle \end{bmatrix}$

Baččeha

mitavânand

[2] $\begin{bmatrix} S \\ \text{XARG} & \langle[1]\rangle \\ \text{VAL} & \langle[1]\rangle \\ \text{MARKING} & \textit{none} \end{bmatrix}$

H

[3]NP
$\begin{bmatrix} V \\ \text{XARG} & \langle[1]\rangle \\ \text{VAL} & \langle[1],[3]\rangle \\ \text{ARG-ST} & \langle[1],[3]\rangle \end{bmatrix}$

madrase

beravand
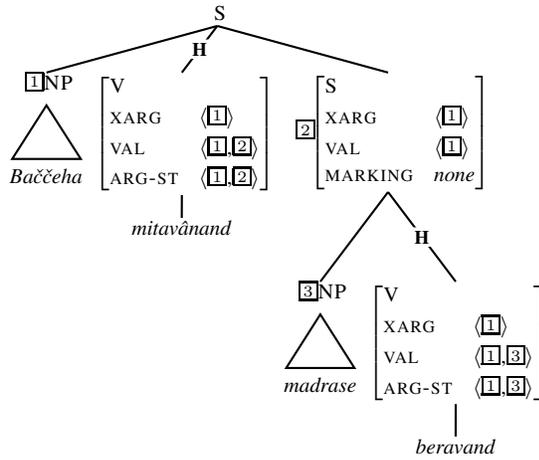
Figure 7
A raising verb with a personal complement (13a)

Finally, raising constructions are licensed by lexical entries such as (34). The crucial difference between this and previous lexical entries is that the verb identifies its XARG with that of its complement. As a side effect, the raising verb will take a subject just in case its complement takes a subject—thus it may appropriately combine with impersonal verbs (15). Figures 7 and 8 contrast the respective analyses of (13a) and (15). The complement will be an unsaturated phrase if and only if the embedded verb takes a canonical subject.

(34) $\begin{bmatrix} \text{LID} & \textit{tavânestan} \\ \text{ARG-ST} & \boxed{L} \oplus \left\langle \begin{bmatrix} \text{HEAD} & V[\textit{sbjv}] \\ \text{XARG} & \boxed{L} \\ \text{LIGHT} & - \end{bmatrix} \right\rangle \\ \text{XARG} & \boxed{L} \end{bmatrix}$

2.8 Synthetic conjugation in HPSG/PFM

Before we address the analysis of periphrastic forms, we start with an account of synthetic conjugation. Table 1 lists all relevant forms of the verb *xaridan* 'buy'.

Persian verbs exhibit a morphomic stem alternation in the sense of (Aronoff 1994); in Table 1 the alternation is *xar* vs. *xarid*. Although the two stems are traditionally called 'present' and 'past', neither of them is associated with a coherent set of paradigm cells: notice in particular that stem selection can not
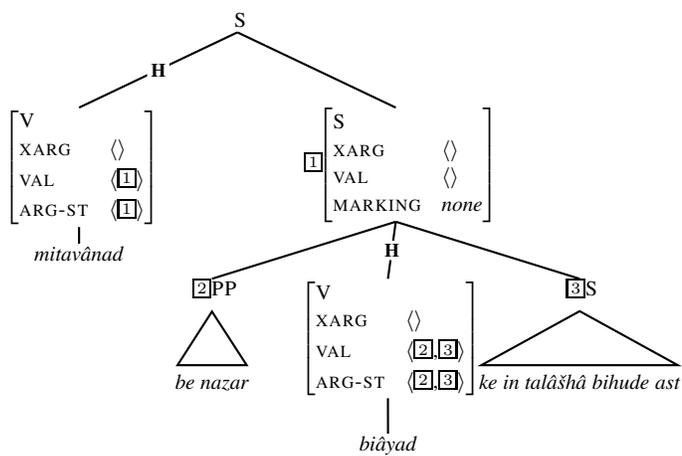
Figure 8
A raising verb with an impersonal complement (15)

	POS	NEG
1.SG	*mixaram*	*nemixaram*
2.SG	*mixari*	*nemixari*
3.SG	*mixarad*	*nemixarad*
1.PL	*mixarim*	*nemixarim*
2.PL	*mixarid*	*nemixarid*
3.PL	*mixarand*	*nemixarand*

Simple present

	POS	NEG
1.SG	*bexaram*	*naxaram*
2.SG	*bexari*	*naxari*
3.SG	*bexarad*	*naxarad*
1.PL	*bexarim*	*naxarim*
2.PL	*bexarid*	*naxarid*
3.PL	*bexarand*	*naxarand*

Simple subjunctive

	POS	NEG
1.SG	*xaridam*	*naxaridam*
2.SG	*xaridi*	*naxaridi*
3.SG	*xarid*	*naxarid*
1.PL	*xaridim*	*naxaridim*
2.PL	*xaridid*	*naxaridid*
3.PL	*xaridand*	*naxaridand*

Simple bounded past

	POS	NEG
1.SG	*mixaridam*	*nemixaridam*
2.SG	*mixaridi*	*nemixaridi*
3.SG	*mixarid*	*nemixarid*
1.PL	*mixaridim*	*nemixaridim*
2.PL	*mixaridid*	*nemixaridid*
3.PL	*mixaridand*	*nemixaridand*

Simple unbounded past

	POS	NEG
IMPERATIVE (2.SG)	*bexar*	*naxar*
INFINITIVE	*xaridan*	*naxaridan*
SHORT INF	*xarid*	—

Other forms

	POS	NEG
PRST PART.	*xarande*	—
PERFECT PART.	*xaride*	*naxaride*
GERUNDIVE	*xarân*	—

Other forms

Table 1
Synthetic forms of *xaridan* 'buy'

17

'present' stem	xar	gozar	kan	zan	šenav	âfarin
'past' stem	xarid	gozašt	kand	zad	šenid	âfarid
translation	'buy'	'pass'	'dig'	'hit'	'hear'	'create'

Table 2
Unpredictable stem alternations

plausibly be based on tense for non-finite forms. In addition, the form of neither stem is phonologically predictable from that of the other, as illustrated by the small sample in Table 2.

Affixal exponents realize unbounded aspect in the indicative present and past (*mi-*), subjunctive or imperative mood (*be-*), negation (*na-* or *ne-*, not illustrated here), and subject agreement for finite forms. Five non-finite forms are distinguished on the basis of stem selection and suffixes. Notice the contrast between the full infinitive, a form used in nominal contexts (e.g. heading subject phrases) and the short infinitive, which heads the rare non-finite complement clauses of the language.

Within Paradigm Function Morphology,[11] we capture the shape of Persian synthetic paradigms by assuming the features and values in table 3 and the cooccurrence restrictions stated in prose in (35).

(35) a. Mood distinctions are available only for finite forms.

     b. Tense distinctions are available only in the indicative mood.

     c. Only finite forms and participles have aspectual distinctions.

     d. There is no bounded present.

     e. Only finite forms exhibit agreement.

     f. The short infinitive and present participle have no negative form.

The rather simple position class system can be accounted for using the system of rule blocks outlined in table 4 and stated in (36–40) using the conventions of Ackerman & Stump (2004). In PFM, realization rules are organized in successive blocks. When attempting to realize a given set of morphosyntactic features, the most specific applicable rule within the block is chosen. For instance, while (40b) asks that finite verbs with a 2SG subject take the suffix *-i*, the more specific (40c) indicates that the suffix is dropped in the imperative.[12]

(36) a. $X_V, \sigma : \{\} \longrightarrow X$'s second stem   (block I)

     b. $X_V, \sigma : \{\text{FORM } fin\} \longrightarrow X$'s first stem

     c. $X_V, \sigma : \{\text{TNS } past\} \longrightarrow X$'s second stem

     d. $X_V, \sigma : \{\text{FORM } ptcp, \text{PRF } -\} \longrightarrow X$'s first stem

feature	values
FORM	*long-inf, short-inf, participle, finite*
MOOD	*indicative, subjunctive, imperative*
TENSE	*present, past*
ASPECT	*bounded, unbounded*
PERSON	*1, 2, 3*
NUMBER	*sg, pl*

Table 3
Features used in the synthetic subparadigm of Persian verbs

III	II	I	IV	V
na-	*mi-*	stem-selection	*-e*	*-am*
ne-			*-ande*	*-i/∅*
be-			*an*	*-ad/∅*
				-im
				-id
				-and

Table 4
Outline of a PFM analysis of Persian synthetic conjugation

(37) $X_V, \sigma : \{\text{MOOD } ind,\ \text{ASP } unbd\} \longrightarrow \textsf{mi}X$ (block II)

(38)   a. $X_V, \sigma : \{\text{POL -}\} \longrightarrow \textsf{na}X$ (block III)

       b. $X_V, \sigma : \{\text{MOOD } ind,\ \text{ASP } unbd,\ \text{POL } -\} \longrightarrow \textsf{ne}X$

       c. $X_V, \sigma : \{\text{MOOD } non\text{-}ind,\ \text{POL } +\} \longrightarrow \textsf{be}X$

(39)   a. $X_V, \sigma : \{\text{FORM } ptcp,\ \text{PRF } +\} \longrightarrow X\textsf{e}$ (block IV)

       b. $X_V, \sigma : \{\text{FORM } ptcp,\ \text{PRF } -\} \longrightarrow X\textsf{ande}$

       c. $X_V, \sigma : \{\text{FORM } inf\} \longrightarrow X\textsf{an}$

(40)   a. $X_V, \sigma : \{\text{PER } 1,\ \text{NB } sg\} \longrightarrow X\textsf{am}$ (block V)

       b. $X_V, \sigma : \{\text{PER } 2,\ \text{NB } sg\} \longrightarrow X\textsf{i}$

       c. $X_V, \sigma : \{\text{PER } 2,\ \text{NB } sg,\ \text{MOOD } imp\} \longrightarrow X$

       d. $X_V, \sigma : \{\text{PER } 3,\ \text{NB } sg,\ \text{TNS } prs\} \longrightarrow X\textsf{ad}$

       e. $X_V, \sigma : \{\text{PER } 1,\ \text{NB } pl\} \longrightarrow X\textsf{im}$

       f. $X_V, \sigma : \{\text{PER } 2,\ \text{NB } pl\} \longrightarrow X\textsf{id}$

       g. $X_V, \sigma : \{\text{PER } 3,\ \text{NB } pl\} \longrightarrow X\textsf{and}$

To illustrate the workings of the system, let us assume that we are trying to produce the negative first person plural bounded past of the verb *xaridan*. The input to the derivation is the pair in (41a), where X (here *xar*) is the verb's basic stem, i (here XARIDAN) is an index uniquely identifying the lexeme at hand, and τ is the set of morphosyntactic features to be realized. We first select the most specific rule in block I. Three rules are applicable: (36a) (which is applicable to any input), (36b), and (36c). The most specific rule happens to be (36c): Because past tense entails finiteness, rule (36b) is less specific. We thus check the lexicon for XARIDAN's second stem, and get as output the pair in (41b). We then go to block II, which contains a single rule realizing unbounded aspect; we thus exit the block with a new prefix (41c). Block III contains two applicable rules expressing negative polarity, the most specific of which, (38b), also expresses unbounded aspect. The prefix *ne-* is thus realized (41d). In PFM, features are not 'consumed' as they are realized, and hence nothing precludes unbounded aspect from being realized twice—in fact it has to be realized each time a rule licenses its realization. Block IV contains no applicable rule: all rules realize a non-finite form. In that situation we exit block IV with no modification. Block V finally contains a single applicable rule, (40e), which results in an appropriate final suffixation.

(41)   a. $\langle X_i, \tau \rangle = \left\langle xar_{\text{XARIDAN}}, \begin{Bmatrix} \text{POL } neg, \text{FORM } \textit{fin}, \text{MOOD } ind, \text{TNS } past, \\ \text{ASP } bnd, \text{PER } 1, \text{NUM } pl \end{Bmatrix} \right\rangle$

       b. $\langle xarid_{\text{XARIDAN}}, \tau \rangle$ (block I)

       c. $\langle mixarid_{\text{XARIDAN}}, \tau \rangle$ (block II)

       d. $\langle nemixarid_{\text{XARIDAN}}, \tau \rangle$ (block III)

       e. $\langle nemixarid_{\text{XARIDAN}}, \tau \rangle$ (block IV)

f. $\langle \textit{nemixaridim}_{\text{XARIDAN}}, \tau \rangle$ (block V)

Since the integration of HPSG and PFM will be essential to our account of periphrastic conjugation, it is important that we specify how we intend to do it. The task is not trivial, because of PFM's reliance on comparisons of feature structure descriptions, which is not readily formulated in existing description languages for HPSG grammars.[13] Rather than attempting a direct integration, we use a PFM grammar to further constrain the class of signs satisfying an HPSG theory. Specifically, we rely on a slight reorganization of the feature geometry for words values as in (42), where MORSYN is an attribute carried by words which collects features that get realized in inflection.[14] The attribute LID is used to assign a specific index to each lexeme (Spencer 2005, Sag 2012).[15] We then define a version of PFM that is exactly like that of Stump (2001) except for the fact that typed feature structures are used to model morphosyntactic property sets. The meta-constraint in (43) then links the two grammars.

(42) $word \rightarrow \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{LID} & \textit{lexemic-index} \\ \text{MORSYN} & \textit{morsyn} \end{bmatrix} \end{bmatrix}$

(43) **Morphology-syntax interface** (provisional version)
A sign of type *word* meeting the description

$$\begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{HEAD} & \begin{bmatrix} \text{LID} & \boxed{2} \\ \text{MORSYN} & \boxed{3} \end{bmatrix} \end{bmatrix}$$

is well-formed only if the PFM grammar licenses phonology $\boxed{1}$ as a realization of the features $\boxed{3}$ for the lexeme $\boxed{2}$.

3. THE PASSIVE

The passive in Persian is a typical complex predicate construction, whose properties are parallel to those of copula-predicative complement constructions (Moyne 1974, Dabir-Moghaddam 1982). The auxiliary *šodan* is clearly the head: all inflectional information, e.g. negation (44), is realized on the auxiliary. The participle-auxiliary sequence is syntactically flexible: the auxiliary can have wide scope over a coordination of participles (45), adverbs may intervene (46), the relative order is not rigid (47), and long-distance fronting of the participle is possible (48).

(44) In tâblo foruxte ne-mi-šav-ad.
this painting sell.PFP NEG-UNBD-become.PRS-3SG
'This painting is not sold.'

(45) In   tâblo   (bevasileye do   nâšenâs)   robude   va   foruxte
     this painting (by   two   strangers) steal.PFP and   sell.PFP
     šod.
     become.PST-3SG
     'This painting was stolen and sold by two strangers.'

(46) In   tâblo   foruxte hatman   šod.
     this painting sell.PFP certainly become.PST.3SG
     'This painting was certainly sold.'

(47) In   tâblo   šod   robude   va foruxte.
     this painting become.PST.3SG steal.PFP and sell.PFP
     'It is this painting which was stolen and sold.'

(48) Foruxte fekr   mi-kon-am   [agar in   tâblo   __
     sell.PFP thought UNBD-do.PRS-1SG   if   this painting
     be-šav-ad   (...)].
     SBJV-become.PRS-3SG
     'I think that if this painting is sold (...).'

A notable contrast between Persian and many languages with periphrastic passives is that there is no evidence for the existence of a passive participle. Persian has clauses headed by a participle, but these are always active, as indicated in (49). Note the contrast with the English situation (50).

(49) a. Nâme=râ   xânde,   Maryam birun raft.
         letter=DDO read.PFP Maryam out   go.PST.3SG
         'Having read the letter, Maryam left.'

     b. * bevasileye Maryam   xande,   nâme   Omid=râ   xošhâl
          by   Maryam   read.PFP,   letter Omid=DDO   happy
          kard.
          do.PST.3SG
          (intended) 'Read by Maryam, the letter made Omid happy.'

(50) a. * Read the letter, Paul felt better.
     b.   Read by Paul, the letter sounded depressing.

The existence of (50) is strong evidence for positing that English possesses a passive participle, although it is always syncretic with the (active) past participle (Aronoff 1994: chap. 1). By the same reasoning, the grammaticality pattern in (49) suggests that passive voice can only be expressed constructionally in Persian.

To capture this idea, we rely on an argument composition analysis in the spirit of Hinrichs & Nakazawa (1994) and subsequent work (e.g. Abeillé & Godard 2002, Bouma & van Nood 1998, Chung 1998, Monachesi 1999). Specifically we propose the lexical entry in (51) for the auxiliary lexeme *šodan*, giving rise to analyses such as that in figure 9. The auxiliary verb combines with a lexical ([LIGHT +]) participle rather than a phrase, promotes the participle's second
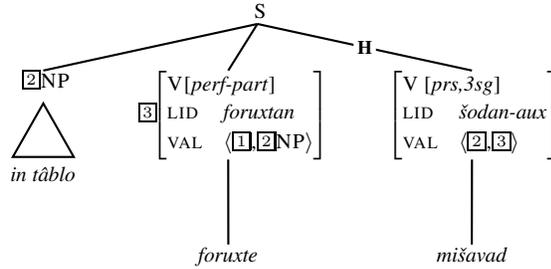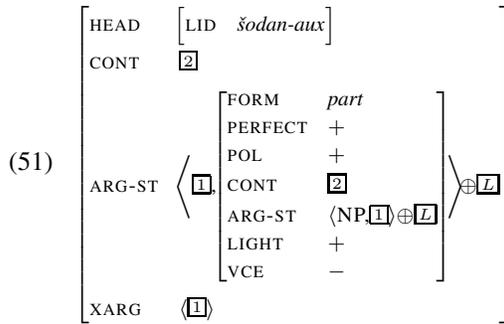
Figure 9
Analysis for (1) 'This painting is sold'

argument (⟦1⟧) to subject status. The logical subject is demoted by not being transmitted to the auxiliary's ARG-ST. All other arguments (⟦L⟧) of the participle are inherited by the auxiliary.[16]

Remember that we distinguish two head-valence schemata for Persian: the *simple head-valence phrase* empties the valence list of its head, save for the subject in some situations (23); while the *verbal complex phrase* forms a small constituent from the head and its least oblique non-subject valent (26). Since the participle in the passive construction is marked as [VCE −], it will combine with the verb and all other valents simultaneously. Since no linear precedence rule intervenes, this will allow the participle to switch places with the auxiliary (47).[17]

To sum up, we propose that voice is not an inflectional category in Persian. The active-passive opposition does not result from two different representations for the main verb, but from the combination of an active participle with an auxiliary verb manipulating its argument structure. The phrase-structural configuration of the passive construction is an independently existing configuration of Persian, analogous to that of the copular construction illustrated in (19) and represented in Fig. 3. As a consequence, passive in Persian is not formally periphrastic inflection.

	POS	NEG		POS	NEG
1.SG	*xaride am*	*naxaride am*	1.SG	*xaride bâšam*	*naxaride bâšam*
2.SG	*xaride i*	*naxaride i*	2.SG	*xaride bâši*	*naxaride bâši*
3.SG	*xaride ast*	*naxaride ast*	3.SG	*xaride bâšad*	*naxaride bâšad*
1.PL	*xaride im*	*naxaride im*	1.PL	*xaride bâšim*	*naxaride bâšim*
2.PL	*xaride id*	*naxaride id*	2.PL	*xaride bâšid*	*naxaride bâšid*
3.PL	*xaride and*	*naxaride and*	3.PL	*xaride bâšand*	*naxaride bâšand*
	Complex present			Complex subjunctive	

	POS	NEG		POS	NEG
1.SG	*xaride budam*	*naxaride budam*	1.SG	*mixaride am*	*nemixaride am*
2.SG	*xaride budi*	*naxaride budi*	2.SG	*mixaride i*	*nemixaride i*
3.SG	*xaride bud*	*naxaride bud*	3.SG	*mixaride ast*	*nemixaride ast*
1.PL	*xaride budim*	*naxaride budim*	1.PL	*mixaride im*	*nemixaride im*
2.PL	*xaride budid*	*naxaride budid*	2.PL	*mixaride id*	*nemixaride id*
3.PL	*xaride budand*	*naxaride budand*	3.PL	*mixaride and*	*nemixaride and*
	Complex bounded past			Complex unbounded past	

	POS	NEG
1.SG	*xaride bude am*	*naxaride bude am*
2.SG	*xaride bude i*	*naxaride bude i*
3.SG	*xaride bude ast*	*naxaride bude ast*
1.PL	*xaride bude im*	*naxaride bude im*
2.PL	*xaride bude id*	*naxaride bude id*
3.PL	*xaride bude and*	*naxaride bude and*
	Complex perfect	

Table 5
The forms based on *budan* for *xaride* 'buy'

4. TWO SETS OF FORMS BASED ON *budan*

There are five different subparadigms based on *budan*, illustrated in Table 5. These contrast in two independent ways.

4.1 *Morphologized vs. truly periphrastic forms*

In the complex present and the complex unbounded past, the perfect participle combines with the present clitic form of the auxiliary, which is homophonous with the exponent of subject agreement except in the 3SG (there is also a non-clitic form of present *budan*, but it may not be used in this construction). In the complex bounded past and complex subjunctive, the perfect participle combines respectively with the bounded past and subjunctive forms of the auxiliary. Finally the complex perfect cumulates two forms of the auxiliary: the participle *bude* and the present form clitic.

There is strong evidence that the forms historically based on the clitic auxiliary have undergone morphologization in contemporary Persian. First, the sequence cannot be interrupted in any way; in particular, adverbs are excluded (52), as is participle fronting (53). Second, colloquial Persian allows a form of vowel reduction in the 3SG that is peculiar to these forms (54a): comparable constructions where the clitic auxiliary combines with a noun or adjective do not give rise to the same pattern (54b). Finally, the distribution of the aspectual marker *mi* is otherwise unexplainable. Since the complex unbounded past indeed realizes unbounded aspect, under a syntactic analysis we would definitely expect the auxiliary to carry the unbounded aspect prefix. Instead, *mi-* is realized before the participle, something which never happens in participial clauses. [18]

(52) a. Hatman rafte=ast.
    certainly leave.PFP=COP.PRS.3SG
    '(S)he has certainly left.'

 b. * Rafte hatman=ast.
    leave.PFP certainly=COP.PRS.3SG

(53) * Ne-mi-rafte sâl-hâ Maryam be madrase=ast.
    NEG-UNBD-leave.PFP year-PL Maryam to school=be.COP.PRS.3SG
    'For years, Maryam didn't go to school'

(54) a. mordˈe=ast → mordˈe:
    die.PFP=COP.PRS.3SG
    '(S)he has died.'

 b. mordˈe=ast → mordˈas
    corpse=COP.PRS.3SG
    'This is a corpse.'

Compare now the situation of forms that are based on a non-clitic auxiliary. The participle-auxiliary combination is more constrained than it is in the passive; in particular, neither adverbs (55) nor object clitics (56) can occur between the two verb forms, and negation must be realized on the participle (57). In addition, the participle rigidly orders before the auxiliary (58). However, the combination is not morphological, since the participle can be extracted (59).

(55) a. Maryam hatman dide bud=aš.
    Maryam certainly see.PFP COP.PST.3SG=CL.3SG
    'Certainly Maryam had seen it.'

 b. * Maryam dide hatman bud=aš.
    Maryam see.PFP certainly COP.PST.3SG=CL.3SG

(56) a. Maryam dide bud=aš.
    Maryam see.PFP COP.PST.3SG=CL.3SG
    'Maryam had seen it.'

25

b. * Maryam dide=aš   bud.
    Maryam see.PFP=3SG COP.PST.3SG

(57) Maryam Omid=râ   na-dide   bud.
    Maryam Omid=DDO NEG-see.PFP COP.PST.3SG
    'Maryam hadn't seen Omid.'

(58) * Maryam Omid=râ   bud   dide.
    Maryam Omid=DDO COP.PST.3SG see.PFP

(59) Foruxte fekr   ne-mi-kon-am   bâš-ad   in
    sell.PFP thought   NEG-UNBD-do.PRS-1SG   COP.SBJV-3SG   this
    tâblo=râ   ].
    painting=DDO
    I don't think that s/he has sold this painting.'

4.2 *Morphosyntactic import*

The use of a form based on *budan* may have two types of semantic import. The complex bounded past (60) and complex subjunctive (61) express respectively the past perfect and the subjunctive perfect. The complex unbounded past however does not express perfectivity at all. Rather, it has an evidential value (Windfuhr 1982, Lazard 1985, Jahani 2000). Whereas the simple unbounded past is used when the speaker has direct evidence for what she is asserting, the complex unbounded past is used in contexts where the evidence is only indirect, as in (62).

(60) Qabl   az   inke Omid be-res-ad,   Maryam birun rafte
    before from that Omid SBJV-arrive.PRS-3SG Maryam out   go.PFP
    bud.
    COP.PST.3SG
    'Maryam had left (before Omid arrived).'

(61) Fekr   mi-kon-am   Maryam mariz bude   baš-ad.
    thought UNBD-do.PRS-1SG Maryam sick   COP.PFP COP.SBJV-3SG
    'I think Maryam has been sick.'

(62) (Banâ bar gofte=ye Omid) Maryam dar sâl=e   1950 in   xâne=râ
    According to=EZ   Omid Maryam in   year=EZ 1950 this house=DDO
    mi-sâxte=ast.
    UNBD-build.PFP=COP.PRS.3SG
    'According to Omid, Maryam would have been building this house in 1950.'

The complex present is ambiguous between a perfect and an evidential value: it can be interpreted either as a present perfect (63a) or as a bounded past with indirect evidentiality (63b). Finally, the complex perfect expresses both perfectivity and indirect evidentiality: it is the indirect evidential equivalent of

	PERFECTIVE	IMPERFECTIVE	PERFECT
PRESENT	***	simple present *mi-xar-ad*	complex present *xarid-e-ast*
PAST DIR.	bounded past *xarid*	unbd. past *mi-xarid*	complex bounded past *xarid-e bud*
PAST IND.	complex present *xarid-e-ast*	complex unbd. past *mi-xarid-e-ast*	complex perfect *xarid-e bud-e-ast*
SUBJUNCTIVE	simple subjunctive *be-xar-ad*		complex subjunctive *xarid-e bâš-ad*

Table 6
The featural content of Persian subparadigms, exemplified in the 3SG

the complex bounded past (64). Note that this corresponds transparently to the fact that the complex perfect includes two realizations of the copula.

(63)   a.   Maryam tâze reside=ast.
            Maryam new arrived=COP.PRS.3SG
            'Maryam has just arrived.'

      b.   (Banâ bar gofte=ye Omid) Maryam in   xâne=râ   dar sâl=e
            According to=EZ   Omid) Maryam this house=DDO in year=EZ
            1950 xaride=ast.
            1950 buy.PFP=COP.PRS.3SG
            'According to Omid, Maryam bought this house in 1950.'

(64)   (Az qarâr), qabl   az   inke Omid be-res-ad,   Maryam birun
      apparently before from that Omid SBJV-arrive.PRS-3SG, Maryam out
      rafte   bude=ast.
      go.PFP COP.PFP=COP.PRS.3SG
      'Apparently, Maryam had left before Omid arrived.'

As can be seen in Table 6, if the present perfect is ignored, morphosyntactic properties align with morphologized vs. syntactic combination: the morphologized forms are used for indirect evidentiality. This can be captured by a simple extension of our morphological analysis: a feature EVID with values *dir* and *indir* is added, subject to the feature cooccurrence restriction in (65). The rules in (66) introduce the relevant exponents. Remember that the clitic copula differs from ordinary subject agreement markers only in the 3SG, so that the correct non-3SG forms are already predicted by the rules in (40)
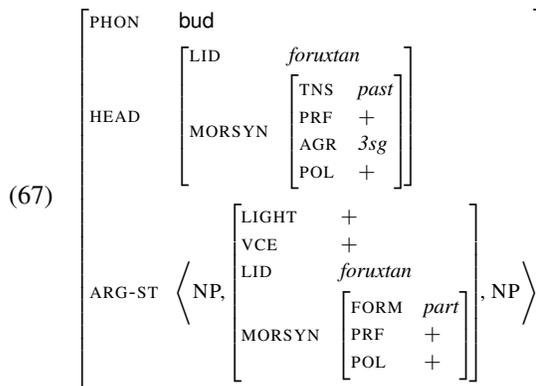
(65)   Evidentiality distinctions are available only in the past.

(66) $X_V, \sigma : \{$EVID *indir*$\} \longrightarrow X$e   (block IV)
      $X_V, \sigma : \{$EVID *indir, 3sg*$\} \longrightarrow X$ast   (block V)

27

The truly periphrastic forms then are used to express the perfect. The fact that the present perfect is unexpectedly synthetic calls for a paradigmatic analysis: this seems to be a standard case of syncretism, where the exponents used to realize a certain feature set (here indirect bounded past) are reused in some unrelated part of the paradigm. However, for such an analysis to be stated, one needs to treat perfect as an inflectional category, and thus to integrate the truly periphrastic forms in the inflectional paradigm (Ackerman & Stump 2004).

4.3 *Exponence as valence*

As we just saw, what we need is a way to treat perfect forms as part of the inflectional paradigm, while allowing for the fact that they correspond to a combination of two words, one of which may be extracted. The solution we explore here can be stated informally as follows: a perfect form of a lexeme Y is a word whose phonology is borrowed from that of a form of the lexeme *budan*, but which subcategorizes for a perfect participle of this same lexeme Y. For instance, the 3SG positive complex bounded past of *foruxtan* meets the description in (67). Notice the discrepancy between the lexemic index and the phonology. As shown in figure 10, because of the VCE + specification, (67) gives rise to a verbal complex, which contrasts with our analysis of the passive (see Fig. 9) and accounts for the tighter solidarity of the verbal sequence. In addition, it will allow for the extraction of the participle through the usual mechanisms for extraction of arguments.

$$
(67) \quad
\begin{bmatrix}
\text{PHON} & \text{bud} \\[4pt]
\text{HEAD} &
\begin{bmatrix}
\text{LID} & \textit{foruxtan} \\[4pt]
\text{MORSYN} &
\begin{bmatrix}
\text{TNS} & \textit{past} \\
\text{PRF} & + \\
\text{AGR} & \textit{3sg} \\
\text{POL} & +
\end{bmatrix}
\end{bmatrix} \\[30pt]
\text{ARG-ST} & \left\langle \text{NP,}
\begin{bmatrix}
\text{LIGHT} & + \\
\text{VCE} & + \\
\text{LID} & \textit{foruxtan} \\[4pt]
\text{MORSYN} &
\begin{bmatrix}
\text{FORM} & \textit{part} \\
\text{PRF} & + \\
\text{POL} & +
\end{bmatrix}
\end{bmatrix}, \text{NP} \right\rangle
\end{bmatrix}
$$

The challenge now is to derive (67) in a principled way, while integrating it within an inflectional system where perfect forms may be realized either synthetically or periphrastically. The approach we propose is based on an extension of the power of realization rules in the spirit of Spencer (2005). In classical PFM, realization rules relate phonology-lexemic index pairs to phonology-lexemic index pairs. We propose that argument lists be added to the picture. Morphosyntactic
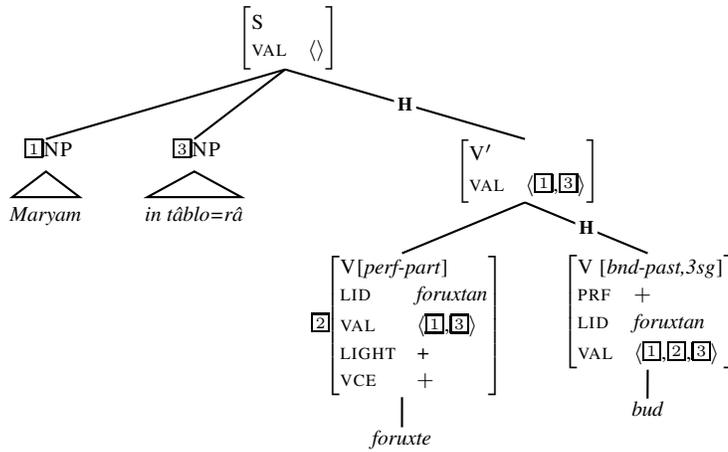
Figure 10
Analysis for (2a) 'Maryam had sold this painting'

features can be realized not only as phonology, but also as valence requirements. We thus extend the meta-constraint in (43) as in (68).

(68) **Morphology-syntax interface** (final version)
A sign of type *word* meeting the description

$$\begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{ARG-ST} & \boxed{2} \\ \text{HEAD} & \begin{bmatrix} \text{LID} & \boxed{3} \\ \text{MORSYN} & \boxed{4} \end{bmatrix} \end{bmatrix}$$

is well-formed only if the PFM grammar licenses phonology $\boxed{1}$ and argument list $\boxed{2}$ as a realization of the features $\boxed{4}$ for the lexeme $\boxed{3}$.

Realization rules must be modified accordingly to include a specification of argument lists. For most rules the change is trivial, as no modification is made to the argument list. The rule licensing (67) is given in (69). This is a portmanteau rule of referral covering blocks I to V, thus bypassing completely synthetic exponence. The rule states that the realization of a morphosyntactic feature bundle σ including a specification for perfect on a lexeme $\boxed{2}$ takes the form of a word with three characteristics. First, its phonology is referred (through the function refer) to that of the corresponding bounded positive non-perfect form of *budan*. The notation '$\sigma \,!\, \tau$' denotes the feature structure that is identical to σ except where for the features mentioned in τ, for which it is identival to τ.[19] Second, this word is an instance of the lexeme $\boxed{2}$, despite having a phonology that does not rely on

one of $\boxed{2}$'s stems. Third, the word carries an extra valence requirement not carried by the lexeme. That requirement is one for a past participle of the lexeme $\boxed{2}$. The participle still carries the subset of features of σ that are compatible with being a participle; in particular it carries the polarity feature, and will thus inflect for negation where appropriate. The meta-constraint in (68) makes sure that the ARG-ST requirement added by the realization rule will indeed constrain the valence of the word within the HPSG grammar. As a consequence, the specification [VCE +] will correctly constrain the participle to form a verbal complex with the participle, unless it is extracted.

(69)
$$
\begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{LID} & \boxed{2} \\ \text{ARG-ST} & \langle\boxed{3}\rangle\oplus\boxed{L} \end{bmatrix}, \sigma : \begin{bmatrix} \text{PRF} & + \end{bmatrix} \longrightarrow
$$

$$
\begin{bmatrix}
\text{PHON} & \text{refer}\left(\begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{LID} & budan \\ \text{ARG-ST} & \langle\boxed{3}\rangle\oplus\boxed{L} \end{bmatrix}, \sigma\,!\,\begin{bmatrix} \text{PRF} & - \\ \text{ASP} & bnd \\ \text{POL} & + \end{bmatrix}, \text{I-V} \right) \\[4ex]
\text{LID} & \boxed{2} \\[1ex]
\text{ARG-ST} & \left\langle \boxed{3}, \begin{bmatrix} \text{LIGHT} & + \\ \text{VCE} & + \\ \text{HEAD} & \begin{bmatrix} \text{LID} & \boxed{2} \\ \text{MORSYN} & \sigma\,!\,\begin{bmatrix} \text{FORM} & part \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \oplus \boxed{L}
\end{bmatrix}
$$

This approach to periphrastic perfect forms has three definite advantages. First, periphrastic inflection is the manifestation of morphosyntactic features in the form of a valence requirement. Thus periphrasis is a variety of syntactic exponence (Blevins to appear), but there is no need to introduce a notion of *phrasal* exponence: the periphrase is licensed by a word-level property. Likewise, no competition between morphology and syntax (e.g. Poser 1992, Bresnan 2001) needs to be orchestrated. Second, since (69) is an inflectional realization rule, it interacts with other such rules under the logic of rule specificity independently needed for affixal exponence. For instance, the fact that the present perfect is syncretic with the (synthetic) indirect bounded past can be accounted for by the rule of referral in (70), which overrides the application of (69) because of specificity.

(70)
$$
\begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{LID} & \boxed{2} \\ \text{ARG-ST} & \boxed{L} \end{bmatrix}, \sigma : \begin{bmatrix} \text{TNS} & prst \\ \text{PRF} & + \end{bmatrix} \longrightarrow
$$

$$
\begin{bmatrix}
\text{PHON} & \text{refer}\left(\begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{LID} & \boxed{2} \\ \text{VAL} & \boxed{L} \end{bmatrix}, \sigma\,!\,\begin{bmatrix} \text{ASP} & bnd \\ \text{PRF} & - \\ \text{EVID} & ind \end{bmatrix}, \text{I-V} \right) \\[4ex]
\text{LID} & \boxed{2} \\[1ex]
\text{ARG-ST} & \boxed{L}
\end{bmatrix}
$$

$$
\begin{bmatrix} \text{S} \\ \text{VAL} \quad \langle\,\rangle \end{bmatrix}
$$

[2]NP

in tâblo

$$
\begin{bmatrix} \text{V}[\textit{perf-part}] \\ \text{LID} \quad \textit{foruxtan} \\ [3]\;\text{VAL} \quad \langle [1],[2] \rangle \\ \text{LIGHT} \quad + \\ \text{VCE} \quad - \end{bmatrix}
$$

foruxte

H

$$
\begin{bmatrix} \text{V}' \\ \text{VAL} \quad \langle [2],[3] \rangle \\ \text{LIGHT} \quad + \end{bmatrix}
$$

$$
[4]\begin{bmatrix} \text{V}[\textit{perf-part}] \\ \text{LID} \quad \textit{šodan-aux} \\ \text{VAL} \quad \langle [2],[3] \rangle \\ \text{LIGHT} \quad + \\ \text{VCE} \quad + \end{bmatrix}
$$

šode

H

$$
\begin{bmatrix} \text{V}\,[\textit{bnd-past,3sg}] \\ \text{PRF} \quad + \\ \text{LID} \quad \textit{šodan-aux} \\ \text{VAL} \quad \langle [2],[4],[3] \rangle \\ \text{LIGHT} \quad + \end{bmatrix}
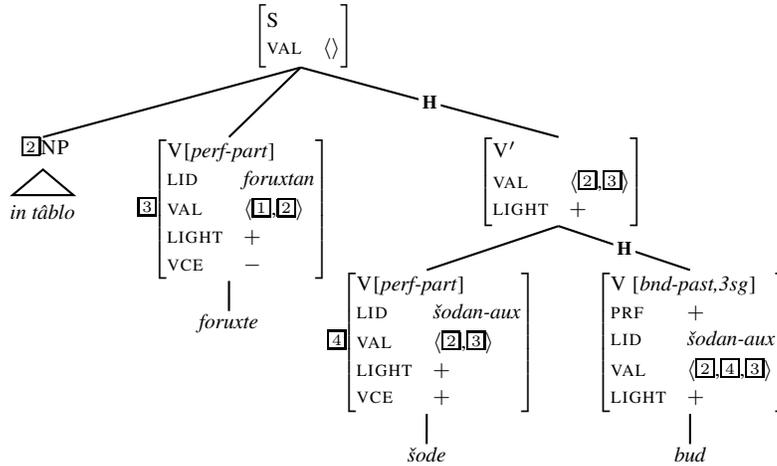$$

bud

Figure 11
A passive version of (2a) 'This painting had been sold'

Third, the present analysis accounts correctly for the interaction between passive and perfect. In principle, one could imagine two realizations for passive perfects (71): either the perfect auxiliary selects the passive or the other way around. In an analysis where both passive and perfect are morphosyntactic features, there is no expectation as to which of these two strategies will be used by the language. Under our present assumptions however, we correctly predict (71a) to be the attested possibility, whose analysis is stated in Figure 11: there is a passive auxiliary lexeme, and this lexeme has a perfect form, which by (69) is a combination of a form of *budan* with the participle *šode*. On the other hand there is no perfect auxiliary lexeme, and no passive inflectional subparadigm; thus there is no basis for generating a combination such as (71b).

(71) a. In   tâblo   foruxte šod-e   bud.
          This picture sell.PFP become-PFP COP.PST.3SG
          'This picture has been sold.'

     b. * In   tâblo   foruxte bud-e   šod.
           This picture sell.PFP COP-PFP become.PST.3SG

4.4 *Discussion*

This general strategy for addressing the Persian perfect, initially proposed in Bonami & Samvelian (2009), has been successfully applied to analogous data in Pamir languages (Stump & Hippisley 2011) as well as other inflectional

periphrasis phenomena in Sanskrit (Stump 2013), Bulgarian (Popova & Spencer 2013), and various Romance and Germanic languages (Bonami & Webelhuth 2011, 2013). The particular formulation presented here raises one potential concern:[20] by modifying valence, rules of periphrasis as stated here are not strictly realisational in the sense of (Stump 2001: 4), as the output of a rule manifests syntactic properties distinct from those of its input. The recent literature has presented two attempts to address this concern, neither of which is fully satisfactory. Stump & Hippisley (2011) introduce valence as a dimension of characterization of paradigm cells. The effect of rule (69) are implemented as the combination of a condition on well-formed paradigms (making room for periphrastic forms with extended valence) and a rule of referral (making explicit the form of the head of a periphrastic construction). While technically this does away with concerns of incrementality, it is clearly unsatisfactory to model a single rule of periphrasis through two separate statements whose independent existence is empirically unmotivated; moreover, it is unclear that the proposal can scale up to cases of stacked periphrases (situations where two auxiliaries simultaneously combine with a main verb to realize morphosyntactic features). Bonami & Webelhuth (2013) propose another alternative, also adopted by Stump (2013). In their analysis, rules of periphrasis do not compete with inflectional realization rules, but with synthetic inflection as a whole; thus the inflection system proper is decidedly realizational. The downside of such a setup is that there can be no direct competition between a realization rule and a rule of periphrasis, making it difficult to state naturally the situation illustrated by the Persian present perfect. Bonami & Webelhuth (2011) sketches a novel solution which reverses the perspective, by taking the main verb to be the principal realization of the lexeme, with the auxiliary realizing features through a process of syntactic reverse-selection. Until this proposal has been fully worked out, we submit that the formal implementation of inflectional periphrasis presented here is still the most adequate to be available.

5. THE FUTURE

The future is formed by combining a special form of the auxiliary *xâstan* 'want' with a short infinitive of the main verb, as indicated in Table 7. Interestingly, there is no compelling argument in favor of a morphological or syntactic analysis.

As in the case of the periphrastic perfect, the verb sequence cannot be interrupted, and occurs in a rigid order (72).

(72) a. Maryam Omid=râ xâh-ad did.
Maryam Omid=DDO want.PRS-3SG see[SINF]
'Maryam will see Omid.'

b. * Maryam xâh-ad Omid=râ did.
Maryam want.PRS-3SG Omid=DDO see[SINF]

c. * Maryam Omid=râ did xâh-ad
Maryam Omid=DDO see[SINF] can.PRS-3SG

	POS	NEG
1.SG	*xâham xarid*	*naxâham xarid*
2.SG	*xâhi xarid*	*naxâhi xarid*
3.SG	*xâhad xarid*	*naxâhad xarid*
1.PL	*xâhim xarid*	*naxâhim xarid*
2.PL	*xâhid xarid*	*naxâhid xarid*
3.PL	*xâhand xarid*	*naxâhand xarid*

Table 7
The future subparadigm of *xaridan* 'buy'

In addition however, unlike what happens in the perfect, no major syntactic rule manipulates the auxiliary-main verb sequence: the main verb can't be fronted (73), nor can a coordination of main verbs combine with an auxiliary (74).

(73) * Did   Maryam Omid=râ   xâh-ad.
       see[SINF] Maryam Omid=DDO want.PRS-3SG

(74) a. Maryam Omid=râ   xâh-ad   did   va xâhad
          Maryam Omid=DDO want.PRS-3SG see[SINF] and want.PRS-3SG
          šenâxt.
          recognize[SINF]
          'Maryam will see and will recognize Omid.'

      b. * Maryam   Omid=râ   xâh-ad   did   va
          Maryam   Omid=DDO   want.PRS-3SG see[SINF]   and
          šenâxt.
          recognize[SINF]

The examples so far show that there is no compelling reason for treating the auxiliary and the main verb as two separate syntactic atoms, as they are always adjacent. The only apparent exception is that object pronominal clitics can optionally be realized in the middle of the sequence (75b). If however the clitics themselves are analyzed as pronominal affixes rather than syntactic atoms (Samvelian & Tseng 2010), this does not preclude a morphological analysis of the combination.

(75) a. Maryam xâh-ad   did=aš.
          Maryam want.PRS-3SG see[SINF]=CL.3SG
          'Maryam will see him.'

      b. Maryam xâh-ad=aš   did.
          Maryam want.PRS-3SG=CL.3SG see[SINF]
          'Maryam will see him.'

Let us now turn to the opposite question: are there compelling reasons to treat the auxiliary-main verb sequence as a single word? None of the arguments

invoked in the case of the complex present carries over. The periphrastic future does not enter into paradigmatic competition with synthetic inflection, and does not give rise to lexically restricted morphophonological idiosyncrasies. The forms composing the periphrase are a bit unusual, but do occur elsewhere in Persian grammar. The conclusion then is that there is little empirical synchronic evidence helping us decide whether we are dealing with one or two syntactic atoms. Nor is there any clear rationale for choosing in terms of overall elegance or simplicity of the analysis: if the auxiliary+main verb sequence is a word, it is an odd kind of compound allowing for infixation of pronominal clitics; if it is a phrase, it is a unique case in the language of a left-headed verbal complex.

In this situation, we turn to diachronic evidence to choose between the two alternatives. Arguably, the analysis positing the smallest departure from the historical source of the construction is preferable. Historically, the future periphrase is a remnant of a complementation option for modal verbs that barely survives in the contemporary language, where the modal combines with a short infinitive (Lazard 1963). (76) presents a relevant attested example taken from Lenepveu-Hotz (2010).

(76) īn pisar-rā maqām va xāna=i harām bi-x^{w}āh-am
this son=DDO sanctuary and house=EZ forbidden FOC-want.PRS-1SG
namūd
show[SINF]
'I want to show to this son the sanctuary and the Ka'aba.'

(*Tārīx-i Sīstān*, 11th century)

In the contemporary language this construction was specialized to two specific contexts: the future periphrase, and the combination of a modal verb with an impersonal complement (77). Notice that in the impersonal complementation construction, unlike what happens in the future periphrase, the short infinitive does not have to be adjacent with the governing verb.

(77) a. Maryam (hatman) bây-ad be madrasa be-rav-ad.
Maryam certainly must.PRS-3SG to school SBJV-go.PRS-3SG
'Maryam definitely has to go to school.'

b. (Hatman) bây-ad be madrasa raft.
certainly must.PRS-3SG to school go[SINF]
'It is definitely necessary to go to school.'

Lenepveu-Hotz (2010) argues that, in combination with the verb *xâstan* 'want', the short infinitive complementation progressively specialized for the expression of the future, while the finite complementation specialized for the expression of volition. Such a tendency is already attested in the 11th century (78). Notice that the use of the aspectual prefix *mi—* with the verb *xâstan* is a later development: at the time, *xâstan* was in line with other modal verbs in not taking the aspectual prefix, a property that few verbs kept in the contemporary language.

(78) a. x^{w}āh-ad   ki   adab   āmōz-ad   ba āsānī
      want.PRS-3SG COMP education learn.PRS-3SG at easyness
      'He wants to be educated without any effort.'

(Tārīx-i Sīstān, 11th century)

    b. man šahr-ē   banā   x^{w}āh-am   kard   [. . .]
      1SG city-IND construction want.PRS-1SG do[SINF]
      'I will build a city. . . '

(Tārīx-i Sīstān, 11th century)

Thus it seems that the most likely analysis for the current state of the language posits that the periphrase involves the syntactic combination of two separate words forming an unusual, left-headed verbal complex. Like the perfect and unlike the passive, this periphrase is integrated in the inflection system: (i) It expresses a morphosyntactic feature that is otherwise expressed synthetically; (ii) It is unexpectedly incompatible with the expression of another morphosyntactic feature, the perfect; (iii) The auxiliary does not have the inflectional characteristics of the main verb it stems from.

Concretely, we list *future* as another possible value of TENSE,[21] and posit the rule in (79). This is similar to the rule for the perfect, with three crucial differences: (i) the phonology of the auxiliary verb is obtained not from referring to the full inflection of a lexeme, but only to the output of rule blocks III to V, applied to the specific string xâh. Thus we avoid positing a second lexeme *xâstan* with a paradigm different from that of the main verb *xâstan*, while still capturing the fact that the auxiliary inflects for polarity and subject agreement; (ii) The main verb is a short infinitive, not a participle; as such it can be subjected to different linear precedence constraints than those affecting the main verb in the perfect periphrase, capturing the difference in word order; Finally (iii) polarity is realized on the auxiliary, not on the main verb.

$$(79) \begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{LID} & \boxed{2} \\ \text{ARG-ST} & \langle\boxed{3}\rangle\oplus\boxed{L} \end{bmatrix}, \sigma: \begin{bmatrix} \text{TNS} & \textit{fut} \end{bmatrix} \longrightarrow$$

$$\begin{bmatrix} \text{PHON} & \text{refer}\left(\begin{bmatrix} \text{PHON} & \text{xâh} \\ \text{LID} & \boxed{2} \\ \text{ARG-ST} & \boxed{L} \end{bmatrix}, \sigma, \text{III-V}\right) \\ \text{LID} & \boxed{2} \\ \text{ARG-ST} & \left\langle \boxed{3}, \begin{bmatrix} \text{LEX} & + \\ \text{VCE} & + \\ \text{HEAD} & \begin{bmatrix} \text{LID} & \boxed{2} \\ \text{MORSYN} & \sigma\,!\begin{bmatrix} \text{FORM} & \textit{short-inf} \\ \text{POL} & + \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle\oplus\boxed{L} \end{bmatrix}$$

This rule correctly accounts for the interaction between the future and other constructions. Because the passive auxiliary is an ordinary lexeme, it has a
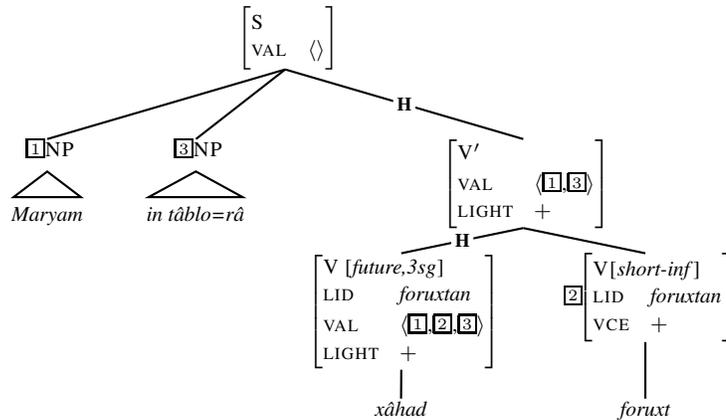
35

$$
\begin{bmatrix} S \\ \text{VAL} \quad \langle\rangle \end{bmatrix}
$$

[1]NP   [3]NP   **H**

Maryam   *in tâblo=râ*

$$
\begin{bmatrix} V' \\ \text{VAL} \quad \langle[1],[3]\rangle \\ \text{LIGHT} \quad + \end{bmatrix}
$$

H

$$
\begin{bmatrix} V\,[\textit{future,3sg}] \\ \text{LID} \quad \textit{foruxtan} \\ \text{VAL} \quad \langle[1],[2],[3]\rangle \\ \text{LIGHT} \quad + \end{bmatrix}
\qquad
[2]\begin{bmatrix} V[\textit{short-inf}] \\ \text{LID} \quad \textit{foruxtan} \\ \text{VCE} \quad + \end{bmatrix}
$$

xâhad   *foruxt*

Figure 12
Analysis for (2a) 'Maryam will sell this painting'

periphrastic future form that is the output of rule (79). However, since the future auxiliary forms a verbal complex with its complement, and the participle in the passive construction is not a verbal complex element, the participle can not be linearized between the two parts of the future periphrase.

(80) a. In tâblo   foruxte [xâh-ad   šod].
           this painting sell.PFP   want.PRS-3SG become[SINF]
           'This painting will be sold.'

      b. * In tâblo   xâh-ad   foruxte šod.
           this painting want.PRS-3SG sell.PFP become[SINF]

When applied to a particle verb, the rule adds a selectional requirement for a short infinitive to a lexical entry already selecting a particle as a verbal complex element. Because that auxiliary, rather than the short infinitive, is the syntactic head of the construction, both the particle and the short infinitive will be realized as complements of the auxiliary. This accounts for the fact that the particle readily linearizes before the auxiliary.

(81) a. Omid ketâb=râ   bar   xâh-ad   dâšt.
           Omid book=DDO PART want.PRS-3SG have[SINF]

      b. Omid ketâb=râ   xâh-ad   bar   dâšt.
           Omid book=DDO want.PRS-3SG PART have[SINF]
      'Omid will take the book.'

Finally, this rule combines readily with the analysis of pronominal object clitics as pronominal affixes defended by Samvelian & Tseng (2010) to account for the ordering possibilities noted in (75).

6. THE PROGRESSIVE

6.1 *The progressive as a periphrase in the typological sense*

All Persian forms compatible with unbounded aspect may give rise to either a progressive or a habitual interpretation when combined with an eventive lexeme. The progressive interpretation can also be forced by using the construction illustrated in (82), based on the verb *dâštan* combined with a finite complement. Notice that the construction is compatible with all and only those forms that are explicitly marked for unbounded aspect—that is, the indicative present (82a), unbounded past (82b), and complex unbounded past (82c), but not e.g. the bounded past (82d) or the subjunctive present (82e)[22].

(82) a.   Maryam dâr-ad   tâblo=râ   mi-foruš-ad.
             Maryam have.PRS-3SG painting=DDO UNBD-sell.PRS-3SG
             'Maryam is selling the painting.'

      b.   Maryam dâšt   tâblo=râ   mi-foruxt.
             Maryam have.PST.3SG painting=DDO UNBD-sell.PST.3SG
             'Maryam was selling the painting.'

      c.   Maryam   dâšte=ast   tâblo=râ
             Maryam   have.PFP=COP.PRS.3SG   painting=DDO
             mi-foruxte=ast.
             UNBD-sell.PFP=COP.PRS.3SG
             'Reportedly, Maryam was selling the painting.'

      d. * Maryam dâšt   tâblo=râ   foruxt.
             Maryam have.PST.3SG painting=DDO UNBD-sell.PST.3SG
             (intended) 'Maryam was selling the painting.'

      e. * Maryam   dâšte   bâš-ad   tâblo=râ
             Maryam   have.PFP   COP.SBJV-3SG   painting=DDO
             be-foruš-ad.
             SBJV-sell.PRS-3SG
             'that Maryam be selling the painting.'

In addition to being unavailable in the subjunctive and bounded past, the progressive periphrase is incompatible with the future (83), the perfect (84), and with the expression of negation (85).

(83) a. * Maryam xâh-ad dâšt   xâh-ad david.
             Maryam FUT-3SG have[SINF] FUT-3SG run[SINF]
             (intended) 'Maryam will be running.'

      b. * Maryam xâh-ad   dâšt   david.
              Maryam FUT-3SG have[SINF] run[SINF]

      c. * Maryam dâr-ad   xâh-ad david.
             Maryam have.PRS-3S FUT-3SG run[SINF]

37

(84) a. * Maryam hatman   dâšte-ast   davide-ast.
          Maryam certainly have.PRS.PRF-3SG run.PRS.PRF-3SG
          (intended) 'Maryam must have been running.'

      b. * Maryam hatman   dâr-ad   davide-ast.
          Maryam certainly have.PRS-3SG run.PRS.PRF-3SG

      c. * Maryam hatman   dâšte-ast   dav-ad.
          Maryam certainly havePRS.PRF-3SG run.PRS-3SG

(85) a. * Maryam na-dâr-ad   (ne-)mi-dav-ad.
          Maryam NEG-have.PRS-3SG NEG-UNBD-run.PRS-3SG
          (intended) 'Maryam is not running.'

      b. * Maryam (na-)dâr-ad   ne-mi-dav-ad.
          Maryam NEG-have.PRS-3SG NEG-UNBD-run.PRS-3SG

It should be noted that except for the incompatibility with bounded aspect, these restrictions on the use of the progressive construction are morphosyntactic rather than semantic in nature. Semantically, there is nothing incoherent with describing e.g. future events in progress. To show this, one may contrast the progressive periphrase with the copular constructions *dar hâle budan* (lit. 'be in the mood for') and *mašqule budan* (lit. 'be occupied by') which express progressivity by lexical means, and which exhibit none of the restrictions just observed. Contrast (82e) with (86), (84) with (87), (83) with (88), and (85) with (89).

(86) Fekr   mi-kon-am   ke Maryam mašqul=e   david-an
      thought UNBD-do.PRS-1.SG that Maryam occupied=EZ run-INF
      bâš-ad.
      be.SUBJ-3.SG
      'I think that Maryam is running.'

(87) Maryam hatman   mašqul=e   david-an bude-ast.
      Maryam certainly occupied=EZ run-INF be.PRS.PRF-3SG
      'Maryam must have been running.'

(88) Maryam mašqul=e   david-an xâh-ad   bud.
      Maryam occupied=EZ run-INF FUT-3SG COP.PST.3SG
      'Maryam will be running.'

(89) Maryam mašqul=e   david-an nist.
      Maryam occupied=EZ run-INF NEG.COP.PRS.3SG
      'Maryam is not running.'

The contrasts between the progressive construction based on *dâštan* and the predicative use of the adjective *mašqul* show that the progressive construction based on *dâštan* expresses a morphosyntactic feature, whose distribution is subject to semantically unmotivated restrictions, rather than simply conveying progressive semantics. This observation puts the construction at hand in the family of inflectional periphrases in the typological sense, as defined in the introduction to this article.

6.2 *Syntactic properties of the construction*

Unlike the periphrastic constructions discussed so far, the progressive periphrase results from the grammaticalization of a finite complement clause construction, and all relevant evidence points to the fact that an embedded clausal structure is still present. [23] The non-auxiliary verb is unmistakably a finite form; it occurs on the right of the auxiliary, as finite complement clauses occur on the right of their head. No complementizer can be used (90a), but since complementizers are not obligatory in Persian, this does not preclude a clausal analysis (90b). Complements normally occur between the two verbs; they can scramble to the left of the auxiliary (91a), but this is also possible with clausal complements (91b). Finally, object clitic pronouns must be realized on the non-auxiliary verb (92a), and cannot climb to the auxiliary (92b).

(90) a. Maryam dâr-ad   (* ke)   ketâb mi-xân-ad.
Maryam have.PRS-3SG   COMP book UNBD-read.PRS-3SG
'Maryam is reading a book.'

b. Maryam mi-xâh-ad   (ke) bâ Omid har ruz be
Maryam UNBD-want.PRS-3SG COMP with Omid every day to
sinemâ be-rav-ad.
theatre SBJV-go.PRS-3SG
'Maryam wants to go to theatre with Omid everyday.'

(91) a. Maryam in ketâb=râ dâr-ad   mi-xân-ad.
Maryam this book=DDO have.PRS-3SG UNBD-read.PRS-3SG
'Maryam is reading this book.'

b. Maryam bâ Omid mi-xâh-ad   (ke) har ruz be
Maryam with Omid UNBD-want.PRS-3SG COMP every day to
sinemâ be-rav-ad.
theatre SBJV-go.PRS-3SG
'Maryam wants to go to theatre with Omid everyday.'

(92) a. Maryam dâr-ad   mi-xân-ad=aš.
Maryam have.PRS-3SG UNBD-read.PRS-3SG-=CL.3SG
'Maryam is reading it.'

b. * Maryam dâr-ad=aš   mi-xân-ad.
Maryam have.PRS-3SG=CL.3SG UNBD-read.PRS-3SG

Finally, we observe that the progressive auxiliary has the characteristic property of a raising verb: Like the raising verb *tavânestan* discussed in section 2.3, it can take an impersonal clause as its complement (93).

(93) Dâr-ad be nazar mi-ây-ad ke in talâsh-hâ
have.PRS-3SG to opinion UNBD-come.PRS-3SG that this effort-PL
bihude=ast.
vain=COP.PRS.3SG
'It is becoming apparent that these efforts are vain.'

6.3 *A non-paradigmatic analysis of the progressive periphrase*

The evidence discussed so far suggests that the optimal analysis of the progressive periphrase relies on the combination of a slightly unusual raising verb with its finite complement. Specifically, we assign the auxiliary *dâštan* the lexical entry in (94), a variant of the entry proposed for *tavânestan* in (34). The effects of this entry in a concrete example are illustrated in Fig. 13.

(94)

$$
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{LID} & \textit{dâštan-aux} \\ \text{MORSYN} & \boxed{0}\begin{bmatrix} \text{ASP} & \textit{prog} \end{bmatrix} \end{bmatrix} \\
\text{CONT} & \boxed{1} \\
\text{ARG-ST} & \boxed{L} \oplus \left\langle \begin{bmatrix} \text{MORSYN} & \boxed{0} \\ \text{MARKING} & \textit{none} \\ \text{CONT} & \boxed{1} \\ \text{XARG} & \boxed{L} \\ \text{LIGHT} & - \end{bmatrix} \right\rangle \\
\text{XARG} & \boxed{L}
\end{bmatrix}
$$
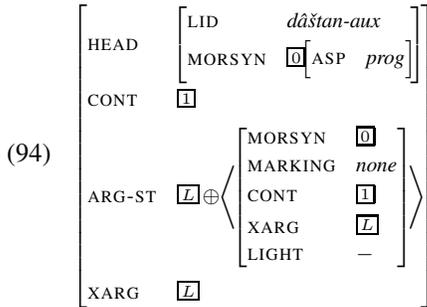
The sharing of XARG value between the auxiliary and its complement captures the auxiliary's subject raising behavior. In the example in Fig. 13, this results in the subject of *miforušad* ($\boxed{1}$) being realized as a valent of *dârad* in the matrix clause. In addition, the auxiliary shares its MORSYN value with that of its complement. As a result, the auxiliary and the main verb must carry the same tense, aspect, mood, polarity, and agreement features. The fact that the progressive is only found in the positive indicative can be captured by feature cooccurrence restrictions: progressive aspect entails unbounded aspect and non-futurity, and aspectuality entails indicative mood. Thus no extra stipulation is needed to make the auxiliary defective for subjunctive, imperative, non-finite forms: this is just a consequence of realizing the progressive feature.

Remember that the periphrase restricts the interpretation of an unbounded form to a progressive interpretation (thus barring e.g. a habitual interpretation), rather than making available an interpretation that is absent when the periphrase is not used. This peculiar situation is captured by making progressive a subtype of unbounded aspect, as indicated in (95).
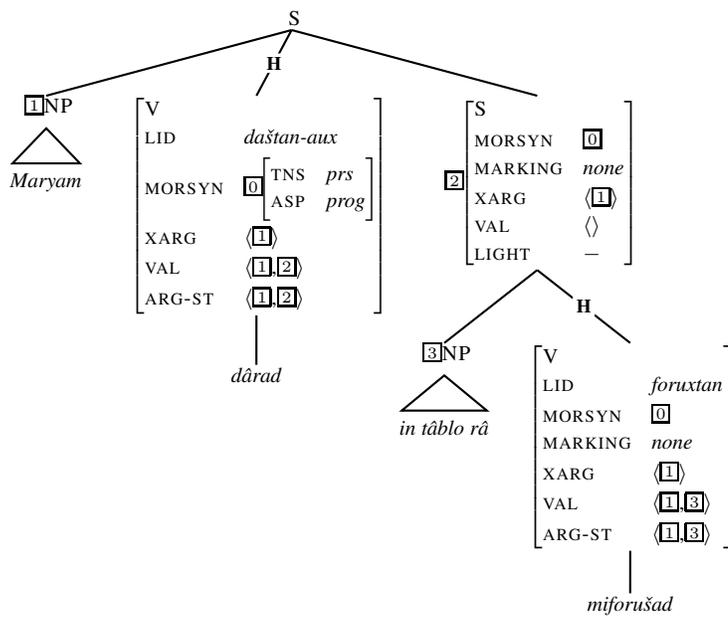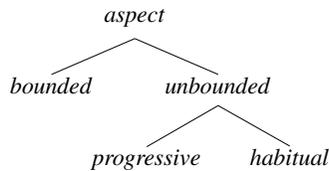
Figure 13
Analysis for (4) 'Maryam is selling this painting.'

(95) Hierarchy of values for the ASPECT feature

aspect

bounded *unbounded*

progressive *habitual*

Because no morphology expresses directly progressivity, unbounded forms (usually marked by the prefix *mi-*) are underspecified between various unbounded interpretations; the periphrase just restricts the interpretation to be progressive. Because of this, the progressive semantics is the realization of a morphosyntactic feature; hence there is no direct semantic contribution to be made by the progressive auxiliary, hence the sharing of CONT values in (94).

7. PERSIAN CONJUGATION AND THE TYPOLOGY OF PERIPHRASIS

With a formally explicit analysis of the five Persian constructions at hand, we are now in a position to examine how they differ and how they fit in a typology of periphrasis. We will discuss three different ways of laying out such a typology: as a scale of morphosyntactic cohesion, as a set of orthogonal categorical criteria, or as a canonical typology.

7.1 *Persian periphrases and degrees of cohesion*

One way of comparing periphrastic constructions is to attempt to place them on a scale of morphosyntactic cohesion, going from open syntactic combination to word-internal stem-affix combinations. Table 8 proposes such a scale. At the bottom is the indirect evidential construction, which we argued to be fully morphologized in contemporary Persian. The future is the limiting case of a syntactic construction with no flexibility at all. The perfect construction is strictly periphrastic: it is a multi-word expression, but very few syntactic processes may affect the sequence; and its distribution is sensitive to paradigm integration. Next comes the progressive construction, which is a specialization of a standard head-finite complement clause combination. Its only distributional peculiarity is that it places a few constraints on the form of the complement clause, which must be subjectless and complementizerless. Finally, the passive construction does not differ at all from a copular construction in terms of cohesion; it is in this sense an analytic combination.

A striking feature of table 8 is that the two constructions we have analyzed as formally periphrastic occupy the center of the scale; from this it is is tempting to infer that formal periphrases in general occupy a middle ground in terms of syntactic cohesion. Yet such a conclusion is unwarranted. Each of the constructions we discussed is in some sense the grammaticalization of a more general

Analytic combination	ordinary head-complement structures	passive
Quasi-analytic	head complement structure, some distributional idiosyncrasies	progressive
True periphrasis	limited syntactic flexibility, paradigm integration	perfect
Quasi-synthetic	no syntactic flexibility two lexemes involved	future
Synthetic combination	ordinary synthetic morphology	evidentials

Table 8

Degrees of periphrasis on a morphosyntactic scale

construction of Persian, and the syntactic properties of the source construction have at least as much to say on cohesion as the formal status of the periphrase (Bonami & Webelhuth 2013). For the passive, progressive and perfect, we argued that the periphrase relies on an independently existing complementation strategy. For the future, there is no other comparable construction in contemporary Persian involving short infinitives that we could rely on for establishing what degree of cohesion is expected. Thus only for evidentials is it clear that the periphrase results in more cohesion than the construction it stems from—a clitic being turned into an affix.

7.2 *Persian periphrases in a multidimensional typological space*

A more appealing way of linking the typology of periphrases and their formal analysis is to apply the criteria for periphrasis collected by Haspelmath (2000) and Ackerman & Stump (2004). We use the formulations of Spencer (2006) for convenience.

(96) a. **Intersectivity**
If a construction expresses grammatical properties that are expressed elsewhere in the synthetic paradigm, then it is periphrastic.

     b. **Noncompositionality**
If some features of elements of the construction are in contradiction with features of the construction as a whole, then the construction is periphrastic.[24]

     c. **Distributed exponence**
If exponence of features of the construction is distributed on the elements of the construction, then the construction is periphrastic.

43

construction	intersective	non-compositional	dist. exp.	underexhaustive
perfect	+	−	+	+
passive	−	+	−	−
progressive	−	−	−	+
future	−	?	−	+

Table 9
Placing Persian periphrases in a typological space

   d. **Underexhaustivity**
If the head of the construction lacks certain forms that other lexemes in the same category have, then the construction is periphrastic.[25]

These four binary criteria define a typological space with 16 possible positions. Notice that these criteria make sense only to distinguish periphrases from open syntactic constructions, and thus they do not apply to the Persian evidential. Table 9 summarizes the classification of the 4 remaining constructions under examination. Two important remarks are in order.

First, if we are to follow Ackerman & Stump (2004) and take the criteria in Table 9 as sufficient reasons for necessitating a formally periphrastic analysis, then it should follow that all four constructions are periphrases since they match at least one criterion. Let us thus review the relevant evidence.

Passive was treated as non-periphrastic, yet it is non-compositional: we have argued that the participle carries a perfect feature that the construction as a whole does not express. Here our analysis takes advantage of the fact that there is no expectation, in any familiar syntactic framework, that features of a non-head element should automatically be features of the construction; if that were the case, non-finite complementation constructions would systematically be treated as periphrases. Hence, we argue, the non-compositionality criterion should be sharpened to the principle in (97).

(97) **Mismatch between morphological exponence and phrasal features**
If the features of a phrasal combination do not match the features that are expected given the exponents carried by the pieces of the construction and independently established general principles of feature transmission, then the combination is periphrastic.

This sharpened principle is a reasonable sufficient condition, under which neither ordinary finite complementation in English nor passive in Persian come out as periphrastic (see also Brown et al. 2012: 252–254).

The second important point in table 9 that is in need of discussion is the fact that the pre-theoretical, or theory-neutral, nature of the criteria should not

be overestimated. A case in point is the compositionality of the future. At first sight it looks non-compositional, because the auxiliary verb looks like a present tense form and the construction as a whole expresses the future. However this is a doubly disputable issue. First, non-compositionality as reformulated in (97) depends on the expression of features. Thus it is crucial to determine that we are indeed dealing with a feature rather than the expression of a semantic predicate by lexical means. Here the evidence discussed in section 6 is relevant: we showed that there is a semantically unmotivated restriction on the cooccurrence between progressive and future, which entails that future must be a feature value participating in the system of paradigmatic oppositions. Second, because there is no synthetic future, there is no way to decide for sure whether the forms of the auxiliary are incompatible with future tense. [26] This situation is unsatisfactory, because if we discard non-compositionality, then there is no remaining distinction in table 9 between the future and the progressive that we can rely on to justify the contrasting analyses proposed in this paper.

What is then the crucial contrast between the future and progressive? The future periphrase is the sole realization of the future, and thus enters a paradigmatic opposition with other values of tense that are expressed synthetically. The progressive periphrase also expresses a feature (progressive aspect), but it is not an obligatory realization of that feature: an imperfective verb, indeed the very same form of used in the periphrase, may have progressive meaning without the presence of the auxiliary. This suggests using obligatoriness as a new criterion:

(98) **Obligatoriness**
If the construction is necessary for some morphosyntactic feature value to be expressed, then the construction is periphrastic.[27]

This makes sense as part of a typology of periphrases, in the sense that we expect periphrases to be obligatory. It also makes sense as a criterion favoring a formally periphrastic analysis: if the inflectional paradigm of a lexeme is defined as the collection of licit combinations of morphosyntactic feature values for that lexeme, a reductionist analysis of an obligatory periphrase is forced to postulate cells in an inflectional paradigm that can not be realized by inflection.

7.3 *Persian periphrases in a canonical typology*

In a recent paper, Brown et al. (2012) outline a different approach to the typology of periphrases based on canonical typology (see e.g. Corbett 2007). The central idea is that a canonical periphrase should correspond to the best possible fit between canonical inflectional morphology and canonical syntax. Although discussing the merits of the proposal is beyond the scope of this paper, we may review the criteria and see how they apply to our Persian examples. The four criteria are listed in (99), paraphrasing (Brown et al. 2012: 245). Notice that these are to be interpreted as criteria for identifying *canonical* periphrases rather

than just periphrases. Thus the fact that some construction one would like to call periphrastic does not meet the criteria should not be taken as refuting them.

(99) a. **Feature realization**
         A canonical periphrastic construction realizes a (canonical) grammatical feature.

      b. **Paradigmaticity**
         A canonical periphrastic construction occupies a cell in an otherwise inflected paradigm.

      c. **Transparency** A canonical periphrastic construction exhibits a transparent relation between meaning and form.

      d. **Functional syntax** A canonical periphrastic construction is a canonical functional syntactic construction.

A few terminological clarifications are in order. First, *grammatical feature* is to be interpreted in the way we have used *morphosyntactic feature*. Second, 'inflected paradigm', in the present context, should be interpreted as a paradigm of synthetically inflected words. Paradigmaticity can be seen as a generalization of intersectivity, taking into account the fact that realizing a feature value that is in paradigmatic opposition to other synthetically expressed values of the same feature may be enough to recognize a paraphrase; hence the Persian future, while not exhibiting intersectivity, exhibits paradigmaticity. The criterion of transparency goes in the opposite direction from non-compositionality (or our revised version, mismatch). This is motivated by the observation that both canonical morphology and canonical syntax are transparent. Finally, functional syntactic constructions are taken to be constructions where grammatical words and/or expression of grammatical meaning plays a crucial role.

The application of the criteria to the Persian data is outlined in Table 10. Notice that we have left undecided whether the future is transparent, just as we left undecided whether it was compositional. We listed the passive as expressing a grammatical feature, assuming a broad typological definition of the notion.

A striking result is that the perfect comes out as closest to the canonical periphrase, and the passive as furthest, which aligns nicely with the analyses proposed in the preceding pages. We do not see such a nice alignment though between the place of the progressive in the typology and the analysis we have proposed.

The reasons for these alignments and misalignments is clear. The typology of is based on the idea that canonical periphrasis is canonical inflection, and that canonical inflection is defined in contrast to derivation (Brown & Hippisley 2012: 239). But even when it is realized synthetically, the status of passive as inflection or derivation is disputed and uncertain (see Walther 2013: chapter 1 and references therein). Thus arguably the passive periphrase is less canonical just because of the feature it expresses. The position of the progressive in the typology is dependent on the use of the criterion of transparency is the opposite

46

construction	grammatical feature	paradigmatic	transparent	functional syntax
perfect	+	+	+	+
passive	+	−	−	+
progressive	+	−	+	+
future	+	+	?	+

Table 10

Persian periphrases in the canonical typology of periphrases

direction from that used in previous work such as Ackerman & Stump (2004). This move however seems debatable. The guiding idea behind the criteria is that canonical periphrasis should be taken as the tightest possible fit between canonical syntax and canonical morphology. Yet in general, criteria in canonical typology are used to contrast two classes of phenomena (affixes vs. words, agreement vs. pronouns, inflection vs. derivation, etc.). Since transparency does not contrast grammatical components or types of constructions, it is usually not used as a criterion, e.g. when defining canonical morphology or canonical inflection—of course, we expect any kind of linguistic combination to be transparent, but that does not help us classify them. Thus while it seems coherent to say that within a canonical approach non-compositionality should not be taken to be a criterial property of periphrasis, we contend that its opposite, transparency, should not either.

If transparency is left aside, paradigmaticity remains as the sole point of contrast between our four constructions within the canonical typology. While it is reassuring that the typological classification coincides with our opposition between formally periphrastic and syntactically reducible periphrases, it is striking that the sole criterion of paradigmaticity is far from doing justice to the various divergences between the constructions we have observed.

8. CONCLUSIONS

In this paper we have proposed analyses for five different constructions which are periphrastic in the typological sense. Of these, two (the passive and the progressive) proved to be best analyzed as instances of ordinary syntax, and one (the indirect evidential) as an instance of morphology.

While the last two constructions, the perfect and the future, are optimally analyzed as formally periphrastic, we have shown that formulating such an analysis did not force one to generate phrases in the inflectional system. By combining the HPSG feature geometry for valence selection and the PFM internal notion of a rule of referral, we were able to treat periphrasis as a word-level

phenomenon, whereby a word borrows the morphological realization of an auxiliary verb while selecting for a participial form of the lexeme it realizes. Thus no competition between morphology and syntax (Poser 1992, Bresnan 2001, Kiparsky 2005) needs to be orchestrated, and the inflectional component is not extended beside the description of the properties of words (Blevins to appear). Moreover, as argued in detail by Bonami & Webelhuth (2013), building the theory of inflectional periphrasis on top of the theory of valence makes the correct prediction that periphrases typically exhibit the properties of normal valence-reducing constructions of the language. In the present instance, this is clearly the case for the perfect construction, whose syntactic properties directly follow from the properties of the forms being combined. Interestingly, the future construction provides a contrasting example: as we argued in section 5, the future periphrase fossilized a valence construction that has otherwise been lost by the language.

In this paper we have adopted a conservative approach to the integration of periphrases in the inflectional analysis: wherever both a formally periphrastic and a purely syntactic analysis were defendable, we elected to err in the direction of pure syntax. Our motivation for doing so was twofold: First the use of well-understood and time-proven formal tools is generally preferable, wherever that does not lead to a distortion of the data. Second, by giving purely syntactic analyses their best chance, we have strengthened the case that some constructions are inescapably formally periphrastic. Third, the exercise has forced us to devise analyses that fit the data as closely as possible, and has thus helped sharpen the typology of periphrases.

This is of course not the only possible strategy to approach the topic at hand. The obvious alternative is to develop a formal theory of periphrases that covers as large as possible a subset of the full typology. That strategy is, we think, implicit in work such as Ackerman & Stump (2004), Blevins (to appear), Ackerman et al. (2011), or Bonami & Webelhuth (2013). As it happens, the rule format for periphrastic inflection introduced in section 4.3 could easily be applied to the Persian passive, progressive, and evidential constructions. In each case though, some aspect of the analysis would require a few inelegant stipulations that are avoided under a reductionist analysis. We thus contend that our reductionist strategy has been fruitful: allowing diverse formal encodings for periphrastic constructions allows for a better understanding of the fine properties of these constructions.

More generally, by discussing the place of Persian periphrases in different typological proposals, we showed in section 7 that our current understanding of the typology is quite partial, and can be sharpened by relying on detailed formal analyses which allow one to substantiate, or in some cases, abandon, the criteria on which the typology is based. Thus at this stage in our understanding of the phenomenon of periphrasis, it seems that, in addition to broad surveys such as e.g. Anderson (2006), detailed formal analyses of particular languages such as the one proposed here are indispensable to the construction of a viable typology.

REFERENCES

Abeillé, A., Bonami, O., Godard, D. & Tseng, J. (2006). The syntax of French *à* and *de*: an HPSG analysis. In Saint-Dizier, P. (ed.), *Dimensions of the syntax and semantics of prepositions*, Springer. 147–162.

Abeillé, A. & Godard, D. (2000). French word order and lexical weight. In Borsley, R. D. (ed.), *The nature and function of syntactic categories*, New York: Academic Press, Inc. Syntax and semantics, 325–360.

Abeillé, A. & Godard, D. (2002). The syntactic structure of French auxiliaries. *Language* **78**. 404–452.

Ackerman, F. & Stump, G. T. (2004). Paradigms and periphrastic expression. In Sadler, L. & Spencer, A. (eds.), *Projecting morphology*, Stanford: CSLI Publications. 111–157.

Ackerman, F., Stump, G. T. & Webelhuth, G. (2011). Lexicalism, periphrasis, and implicative morphology. In Borsley, R. D. & Börjars, K. (eds.), *Non-transformational syntax: Formal and explicit models of grammar*, Oxford: Wiley-Blackwell. 325–358.

Ackerman, F. & Webelhuth, G. (1998). *A theory of predicates*. Stanford: CSLI Publications.

Anderson, G. (2006). *Auxiliary verb constructions*. Oxford: Oxford University Press.

Aronoff, M. (1994). *Morphology by itself*. Cambridge: MIT Press.

Barjasteh, D. (1983). *Morphology, syntax and semantics of Persian compound verbs: A lexical approach*. Ph.D. dissertation, University of Illinois.

Blevins, J. P. (to appear). Periphrasis as syntactic exponence. In Blevins, J. P., Ackerman, F. & Stump, G. T. (eds.), *Paradigms and periphrasis*, Stanford: CSLI Publications.

Bonami, O. & Crysmann, B. (to appear). Morphotactics in an information-based model of realisational morphology. In *Proceedings of HPSG 2013*. Stanford: CSLI Publications.

Bonami, O. & Samvelian, P. (2008). Sorani Kurdish person markers and the typology of agreement. Paper presented at the *Thirteenth International Morphology Meeting*, Vienne.

Bonami, O. & Samvelian, P. (2009). Inflectional periphrasis in Persian. In Müller, S. (ed.), *Proceedings of the HPSG 2009 conference*. Stanford: CSLI Publications. 26–46.

Bonami, O. & Stump, G. T. (to appear). Paradigm Function Morphology. In Spencer, A. (ed.), *Handbook of morphology*, Oxford: Wiley-Blackwell. 2nd edition.

Bonami, O. & Webelhuth, G. (2011). Inflectional periphrasis as collocation. Paper presented at the 8th Mediterranean Morphology Meeting.

Bonami, O. & Webelhuth, G. (2013). The phrase-structural diversity of periphrasis: a lexicalist account. In Chumakina, M. & Corbett, G. G. (eds.), *Periphrasis*, Oxford: British Academy and Oxford University Press. 141–167.

Börjars, K., Vincent, N. & Chapman, C. (1997). Paradigms, periphrases and pronominal inflection: a feature-based account. In Booij, G. & van Marle, J. (eds.), *Yearbook of morphology 1996*, Dordrecht: Kluwer. 155–180.

Bouma, G., Malouf, R. & Sag, I. A. (2001). Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* **19**. 1–65.

Bouma, G. & van Nood, G. (1998). Word order constraints on verb clusters in German and Dutch. In Hinrichs, E., Kathol, A. & Nakazawa, T. (eds.), *Syntax and semantics, vol. 30: Complex predicates in nonderivational syntax*, San Diego: Academic Press. 43–72.

Bresnan, J. (2001). *Lexical-functional syntax*. Oxford: Basil Blackwell.

Brown, D., Chumakina, M., Corbett, G. G., Popova, G. & Spencer, A. (2012). Defining 'periphrasis': key notions. *Morphology* **22**. 233–275.

Brown, D. & Hippisley, A. (2012). *Network morphology: a defaults-based theory of word structure*. Cambridge: Cambridge University Press.

Chung, C. (1998). Argument composition and long-distance scrambling in korean: An extension of the complex predicate hypothesis. In Hinrichs, E., Kathol, A. & Nakazawa, T. (eds.), *Syntax and semantics, vol. 30: Complex predicates in nonderivational syntax*, San Diego: Academic Press. 159–220.

Corbett, G. G. (2007). Canonical typology, suppletion and possible words. *Language* **83**. 8–42.

Crysmann, B. & Bonami, O. (2012). Establishing order in type-based realisational morphology. In Müller, S. (ed.), *Proceedings of HPSG 2012*. Stanford: CSLI publications. 123–143.

Dabir-Moghaddam, M. (1982). Passive in Persian. *Studies in the Linguistic Sciences* **12**. 63–90.

Ghomeshi, J. (1997). Topics in Persian VPs. *Lingua* **102**. 133–167.

Greenberg, J. (1954). A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* **26**. 178–194.

Haspelmath, M. (2000). Periphrasis. In Booij, G., Lehmann, C. & Mugdan, J. (eds.), *Morphology: an international handbook on inflection and word-formation*, Berlin: Walter de Gruyter. vol. 1, 654–664.

Hinrichs, E. & Nakazawa, T. (1994). Linearizing AUXs in German verbal complexes. In Nerbonne, J., Netter, K. & Pollard, C. (eds.), *German in HPSG*, Stanford: CSLI Publications. 11–37.

Hockett, C. F. (1958). *A course in modern linguistics*. Macmillan.

Jahani, C. (2000). Expressions of indirectivity in spoken Modern Persian. In Johanson, L. & Utas, B. (eds.), *Evidentials: Turkic, Iranian and neighbouring languages*, Berlin: Mouton de Gruyter. 185–207.

Karimi, S. (1996). Case and specificity: Persian râ revisited. *Linguistic Analysis* **26**. 74–194.

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

Karimi, S. (2008). Raising and control in Persian. In Karimi, S., Samiian, V. & Stilo, D. (eds.), *Aspects of iranian linguistics*, Newcastle: Cambridge Scholars.

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

Kiparsky, P. (2005). Blocking and periphrasis in inflectional paradigms. In Booij, G. & Marle, J. v. (eds.), *Yearbook of morphology 2004*, Springer. 113–135.

Lazard, G. (1963). *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck.

Lazard, G. (1985). L'inférentiel ou passé distancié en persan. *Studia Iranica* **14**. 27–42.

Lazard, G., Richard, Y., Hechmati, R. & Samvelian, P. (2006). *Grammaire du persan contemporain*. Tehran: IFRI and Farhang-e Moaser.

Lenepveu-Hotz, A. (2010). Des verbes modaux persans et de l'évolution de leur construction entre x^e et xvie siècles. *Studia Iranica* **39**. 79–107.

Miller, P. (1992). *Clitics and constituents in phrase structure grammar*. New York: Garland.

Miller, P. & Sag, I. A. (1997). French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* **15**. 573–639.

Monachesi, P. (1999). *A lexical approach to Italian cliticization*. Stanford: CSLI Publications.

Moyne, A. J. (1974). The so-called passive in Persian. *Foundations of language* **12**. 249–67.

Müller, S. (2002). *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German*. Stanford: CSLI Publications.

Müller, S. (2010). Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* **46**. 601–655.

Pollard, C. & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Stanford: CSLI Publications; The University of Chicago Press.

Popova, G. & Spencer, A. (2013). Relatedness in periphrasis: a paradigm-based perspective. In Chumakina, M. & Corbett, G. G. (eds.), *Periphrasis*, British Academy and Oxford University Press. 191–225.

Poser, W. J. (1992). Blocking of phrasal constructions by lexical items. In Sag, I. A. & Szabolcsi, A. (eds.), *Lexical matters*, Stanford: CSLI Publications. 111–130.

Sadler, L. & Arnold, D. (1994). Prenominal adjectives and the phrasal/lexical distinction. *Journal of Linguistics* **30**. 187–226.

Sadler, L. & Spencer, A. (2001). Syntax as an exponent of morphological features. In Booij, G. & van Marle, J. (eds.), *Yearbook of morphology 2000*, Dordrecht: Kluwer. 71–96.

Sag, I. A. (2012). Sign-based construction grammar: An informal synopsis. In Boas, H. & Sag, I. A. (eds.), *Sign-based construction grammar*, Stanford: CSLI Publications. 69–202.

Samvelian, P. (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, P. (2007). A (phrasal) affix analysis of the persian ezafe. *Journal of Linguistics* **43**. 605–645.

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

Samvelian, P. & Tseng, J. L. (2010). Persian object clitics and the syntax-morphology interface. In Müller, S. (ed.), *Proceedings of the HPSG10 conference*. Stanford: CSLI Publications. 212–232.

Spencer, A. (2005). Inflecting clitics in generalized paradigm function morphology. *Lingue e Linguaggio* **4**. 179–193.

Spencer, A. (2006). Periphrasis. In Brown, K. (ed.), *The encyclopedia of language and linguistics, second edition*, Oxford: Elsevier. 287–294.

Stump, G. T. (2001). *Inflectional morphology. a theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, G. T. (2006). Heteroclisis and paradigm linkage. *Language* **82**. 279–322.

Stump, G. T. (2013). Periphrasis in the Sanskrit verb system. In Chumakina, M. & Corbett, G. G. (eds.), *Periphrasis*, British Academy and Oxford University Press. 105–138.

Stump, G. T. & Hippisley, A. (2011). Valence sensitivity in Pamirian past-tense inflection: a realizational analysis. In Korn, A., Haig, G., Karimi, S. & Samvelian, P. (eds.), *Topics in Iranian linguistics*, Wiesbaden: Ludwig Riechert Verlag. 103–116.

Taleghani, A. H. (2008). *Modality, aspect and negation in Persian*. Amsterdam: John Benjamins.

Tseng, J. (2002). Remarks on marking. In Van Eynde, F., Hellan, L. & Beermann, D. (eds.), *The proceedings of the 8th international conference on head-driven phrase structure grammar*. Stanford: CSLI Publications. 267–283.

Tseng, J. L. (2003). Edge features and French liaison. In Kim, J.-B. & Wechsler, S. (eds.), *The proceedings of the 9th international conference on head-driven phrase structure grammar*. Stanford: CSLI Publications. 313–333.

Walther, G. (2013). *De la canonicité en morphologie: perspective empirique, théorique et computationnelle*. Ph.D. dissertation, Université Paris Diderot.

Windfuhr, G. (1982). The verbal category of inference in Persian. In *Momentum Georg Morgenstierne II*. Leiden: E. G. Brill. 263–287.

FOOTNOTES

1 Parts of the work reported here were presented at the 6th *Décembrettes* conference (Bordeaux, 2008), at the 16th HPSG conference (Göttingen, 2009), at the 3rd International Conference on Iranian Linguistics (Paris, 2009), and at the conference on the Typology of Periphrasis (Guildford, 2010). We thank the audience at these events for their questions, suggestions and helpful disagreement, and in particular Marina Chumakina, Grev Corbett, Nick Evans, Andrew Hippisley, Andrew Spencer, and Greg Stump. For reading through various versions of this manuscript and making many valuable suggestions we thank Farrell Ackerman, Bob Borsley, Agnès Lenepveu-Hotz, Stefan Müller, Gert Webelhuth, and two anonymous reviewers. This work was funded by the ANR-DFG project PERGRAM [grant no. MU 2822/3-I] and benefited from interactions with other members of the project, most importantly Stefan Müller and Jesse Tseng.

2 Arguably, the only other formally explicit theory of morphology with appropriate scope is Network Morphology (Brown & Hippisley 2012), but the logic of paths underlying the DATR language in which Network Morphology is written is very distant from the logic of feature structures.

3 The glosses mostly follow the Leipzig Glossing Rules. The following nonstandard abbreviations are used for clarity: BD: bounded aspect; CL: enclitic object pronoun; COP: copula; DDO: definite direct object; EZ: ezafe; PART: particle; PFP: perfect participle; SINF: short infinitive; UNBD: unbounded aspect.

4 This definition is purposefully resonant with the following passage in (Aronoff 1994: 126):

> Derivation and inflection are not kinds of morphology but rather uses of morphology: inflection is the morphological realisation of syntax, while derivation is the morphological realization of lexeme formation [. . .] and the same morphology can sometimes serve both

We likewise assume that lexeme formation may manifest itself by either morphological or periphrastic means, as (Ackerman & Webelhuth 1998, Ackerman et al. 2011) illustrate in detail. Indeed Persian complex predicates are a particularly clear instance of periphrastic lexeme formation, as argued by Samvelian (2012).

5 The distribution of enclitic object pronouns is intricate, and its description and modeling are quite outside the scope of this paper. While the clitic is most often realized as an enclitic to the verb, it may also realize on a dependent of the verb. Samvelian & Tseng (2010) note that the clitic is realized on the least oblique complement (apart from the one realized by the clitic itself). While

this complement is often adjacent to the verb (i), it need not be, and a complement-clitic sequence can be topicalized (ii). Moreover, when attached to a non-verb, the clitic is best analyzed as a phrasal (right edge) affix, since it is subject to various morphophonological idiosyncrasies such as haplology with a form-identical possessive element (iii).

(i)   Omid be Maryam=aš   dâd.
       Omid to Maryam=CL.3SG give.PST.3SG
       'Omid gave it to Maryam.'

(ii)   be Maryam=aš   fekr   mi-kon-am   ke Omid dâd
       to Maryam=CL.3SG think.PST.3SG UNBD-do.PRS-1SG that Omid give.PST.3SG
       'To Maryam, I think that Omid gave it.'

(iii) * Omid be xâhar=aš=aš   dâd
       Omid to sister=CL.3SG=CL.3SG give.PST.3SG
       (intended) ' Omid gave it to her sister.'

6   Some (but not all) of these predicates are also compatible with a non-finite complement headed by a short infinitive verb, in which case the embedded subject receives an arbitrary interpretation (Karimi 2008).

(i)   Bây-ad   madrase raft.
       must.PRS-3SG school   go.SINF
       'One must go to school.'

7   We assume standard HPSG analytic tools and notation. Attributes are typeset in small caps, and types in italics. Feature structures are shown in square brackets, lists of feature structures in angle brackets; parentheses are only used to denote the scope of operations. '⊕' denotes list concatenation, while '◯' denotes shuffle: the shuffle of two lists $l \bigcirc l'$ denotes any list that contains all the elements in l and all the elements in l', respecting both their ordering in l and their ordering in l', but possibly interspersing elements of l and elements of l'. Boxed tags indicate identity between the value of two features; boxed numbers are used for feature structures while boxed uppercase letters are used for lists.

8   This is a consequence of the use of the 'shuffle' operator ◯ in the specification of valence on the head of the phrase: the order of the branches needs not reflect the order on the head's valence list.

9   The feature LIGHT is analogue to the feature LEX as used e.g. by Sadler & Arnold (1994), Müller (2002, 2010). We avoid the name LEX which is misleading: it is unintuitive to talk of lexical phrases or non-lexical words.

10 See Samvelian & Tseng (2010) for a full analysis of Persian pronominal object clitics compatible with the present proposals.

11 See Bonami & Stump (to appear) for an up-to-date presentation.

12 The distinction between blocks IV and V is motivated by data to be discussed in section 3—as they stand, these two blocks could be fused with no consequence on the forms generated.

13 Crysmann & Bonami (2012), Bonami & Crysmann (to appear) design an inflectional component for HPSG grammars that is very similar in spirit to PFM. However extending that component to the treatment of periphrasis is beyond the scope of the present paper.

14 Features values in MORSYN are normally reentrant with feature values in other parts of the representation: tense, aspect and mood are also head features; agreement features are reentrant with features of arguments on ARG-ST; in languages which register inflectionally extraction, MORSYN will record the presence of an element in SLASH (Bouma et al. 2001); in languages with edge inflection it will include feature values reentrant with EDGE features (Tseng 2003). In most situations, the postulation of MORSYN is just a convenience that allows for a clean formulation of the morphology-syntax interface. However it is crucial to capturing appropriately various types of morphosyntactic mismatches, including default agreement for verbs with no nominal subject, or deponency phenomena. The postulation of MORSYN has effects analogous to Sadler &

Spencer's (2001) distinction between syntactic and morphological features, or Stump's distinction between content paradigms and form paradigms.

15 The LID feature is crucial both to capturing the selection of heads of phrases and to the theory of inflection. Whether it should be individuated in semantic terms, as Sag (2012) assumes, is disputable; we make no claims on this here. There is a family resemblance between LID and the PRED feature in LFG, but also important differences. Among other things, all words carry an LID, irrespective of their grammatical status; and LID does not encode any information pertaining to argument structure.

16 An anonymous reviewer suggests that one could encode subject demotion in a phrasal construction rather than in the auxiliary's lexical entry. See Müller (2010) for a detailed argumentation on why such a phrasal analysis of the Persian passive and related phenomena is ill-advised.

17 If verb modifiers are treated as extended valence (e.g. Bouma et al. 2001), this also allows directly for an adverb to occur between the two verbs (46). If not, the analysis of clauses embodied by (23) should be revised so as to allow adjuncts to be interspersed with valents.

18 The only piece of evidence pointing in the other direction is the possibility for the auxiliary to have wide scope over a coordination of participles (i). However, it can be argued that the auxiliary is an edge inflection, similar to other enclitics such as the indefinite determiner =*i*, (ii), or the Ezâfe, (iii), whose morphological status has been thoroughly discussed by Samvelian (2007). Furthermore, the existence of sublexical coordination in numerous languages calls into question whether this is a strong argument against a morphological analysis.

  (i) Sâ'at-hâ [mi-xânde   va mi-raqside]-and
       hour-PL UNBD-sing.PFP and UNBD-dance.PFP-3PL
       'They would have been singing and dancing for hours.'

  (ii) [mard va zan]=i
       man   and woman=INDF
       'a man and a woman'

  (iii) [mard va zan]=e   irâni
        man   and woman=EZ Iranian
        'the Iranian man and a woman'

19 This notation is borrowed from Sag (2012). Stump (2001) uses 'σ/τ' for the same purpose, but we prefer to avoid using the slash, which already has different conventional meaning in the context of HPSG.

20 We are indebted to Greg Stump for initially raising this issue.

21 There is no future perfect, which can readily be captured by a feature coocurrence restriction.

22 The lexeme *dâštan* has no simple subjunctive form; the complex subjunctive is used instead.

23 For an alternative analysis of this construction as a Serial Verb Construction see Taleghani (2008).

24 Note that here we concentrate on the compositionality of features, not of lexical meaning. In most constructions that are candidate for a periphrastic analysis, and in particular in the three relevant Persian constructions, the head lacks its usual lexical semantic import. It is however difficult to use this as a criterion since it could be treated as a simple case of homonymy rather than non-compositionality.

25 Underexhaustivity as defined in (96d) is clearly too general to be used as a sufficient condition: by definition, any defective lexeme gives rise to underexhaustive structures. We suspect that the expectation that periphrases be underexhaustive might be a consequence of the expectation that they be intersective: if a periphrase is intersective, then by definition it will cover only a subpart of the paradigm.

26 The situation in Czech shows this to be a plausible hypothesis. In this language, the future auxiliary (i) coincides with the future form of the copula (ii), which carries inflectional exponents used in the present by the largest class of regular verbs (iii); these same exponents are also carried by the few imperfective verbs with a synthetic future (iv).

(i) Bud-u / bud-eš / bud-e / bud-eme / bud-ete / bud-ou číst tu knihu.
FUT-1SG FUT-2SG FUT-3SG FUT-1PL FUT-2PL FUT-3PL read.INF this book
'I/you/(s)he/we/you/they will read this book.'

(ii) Bud-u / bud-eš / bud-e / bud-eme / bud-ete / bud-ou
COP.FUT-1SG COP.FUT-2SG COP.FUT-3SG COP.FUT-1PL COP.FUT-2PL COP.FUT-3PL
rád / rád-i
happy[M.SG] / happy-M.PL
'I/you/(s)he/we/you/they will be happy.'

(iii) Čt-u / čt-eš / čt-e / čt-eme / čt-ete / čt-ou tu knihu.
read-1SG read-2SG read-3SG read-1PL read-2PL read-3PL this book
'I/you/(s)he/we/you/they read(s) this book.'

(iv) Půjd-u / půjd-eš / půjd-e / půjd-eme / půjd-ete / půjd-ou domů.
go.FUT-1SG go.FUT-2SG go.FUT-3SG go.FUT-1PL go.FUT-2PL go.FUT-3PL home
'I/you/(s)he/we/you/they will go home.'

27 One appeal of obligatoriness is that it is resonant with a well known expectation about inflectional (as opposed to derivational) morphology (Greenberg 1954) that is known to be at best a sufficient condition.