Inflectional periphrasis in Persian

Olivier Bonami

Université Paris-Sorbonne &
Laboratoire de Linguistique Formelle

Pollet Samvelian

Université Paris 3 Sorbonne nouvelle &
Mondes iranien et indien

Proceedings of the HPSG09 Conference

Georg-August-Universität Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

Modern Persian conjugation makes use of five periphrastic constructions. We contrast the properties of these five constructions and argue that they call for different analyses. We propose contrasting analyses relying on the combination of an HPSG approach to feature geometry and syntactic combination, and an approach to paradigm organization and morphological exponence based on Paradigm Function Morphology. This combination of analytic tools allows us to treat the whole array of periphrastic constructions as lexical in origin—no phrasal construction or multi-word lexical entry of any kind is required.

Grammars of Persian (e.g. Lazard et al., 2006) distinguish five 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, and giving rise to different syntactic and semantic properties. The future is formed with a special present tense form of *xâstan* 'want' followed by a bare stem (3). Finally, the progressive is based on an inflected form of *dâštan* 'have' followed by a finite form (4).[1]

(1) In tâblo foruxte mi-šav-ad.
this painting sold UNBD-become.S1-3SG
'This painting is sold.'

(2) a. Maryam in tâblo=râ foruxte bud.
Maryam this painting=DDO sold be.S2.3SG
'Maryam had sold this painting.'

b. Maryam in tâblo=râ foruxe=ast.
Maryam this painting=DDO sold=be.PRS.3SG
'Maryam has sold this painting.'

(3) Maryam in tâblo=râ xâh-ad foruxt.
Maryam this painting=DDO want.S1-3SG sell.S2
'Maryam will sell the painting'

[†]Aspects of this work have been presented at the *Décembrettes 6 International Morphology Conference* (December, 2008), at the HPSG Seminar at U. Paris Diderot (March, 2009), at a Morphology Meeting in Surrey (April, 2009), and at the HPSG 2009 Conference (Göttingen, July 2009). We thank for their comments and suggestions the audiences at these events, and in particular Anne Abeillé, Gilles Boyé, Dunstan Brown, Patricia Cabredo Hofherr, Greville Corbett, Berthold Crysmann, Gerald Gazdar, Stefan Müller, Ivan A. Sag, Gregory Stump, Jesse Tseng, and Gert Webelhuth. This work was supported by a grant from Agence Nationale de la Recherche and Deutsche Forschungsgemeinschaft to the Franco-German project 'PER-GRAM: Theory and Implementation of a Head-driven Phrase Structure Grammar for Persian'.

[1]The glosses use the following abbreviations. BD: bounded aspect; DDO: definite direct object; EZ: Ezafe; NEG: negation; PAF: pronominal affix; PRF: perfect; PRS: present; PST: past; S1: first stem (a.k.a. the present stem); S2: second stem (a.k.a. the past stem); SBJV: subjunctive; UNBD: unbounded aspect.

(4) Maryam dâr-ad in tâblo=râ mi-foruš-ad.
Maryma have.PRS-3SG this painting=DDO UNBD-sell.S1-3SG
'Maryam is selling the painting.'

The differing properties of these five types of periphrasis stem 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, through truly periphrastic forms that need to be integrated into inflectional paradigms despite being multi-word expressions. We assume that the different properties call for different analyses. We propose five contrasting analyses relying on the combination of an HPSG approach to feature geometry and syntactic combination, and an approach to paradigm organization and morphological exponence based on Paradigm Function Morphology (PFM; Stump, 2001). Interestingly, this combination of analytic tools allows us to treat the whole array of periphrastic constructions as lexical in origin—no phrasal construction or multi-word lexical entry of any kind is required.

1 Synthetic conjugation in HPSG/PFM

Before we address the analysis of periphrastic forms, we start with an account of synthetic conjugation. (5) lists the synthetic subparadigms of the lexeme *xaridan* 'buy', using the positive 2SG form as an illustration.

(5) a. Finite forms:
   i. Simple present: *mi-**xar**-i*
   ii. Simple bounded past: ***xarid**-i*
   iii. Simple unbounded past: *mi-**xarid**-i*
   iv. Simple subjunctive: *be-**xar**-i*
   v. Imperative: *be-**xar***
b. Nonfinite forms:
   i. Infinitive: ***xarid**-an*
   ii. Present participle: ***xar**-ande*
   iii. Perfect participle: ***xarid**-e*
   iv. Gerund: ***xar**-ân*

Persian verbs exhibit a morphomic stem alternation (here *xar* vs. *xarid*). Neither stem is predictable from the other in general, and both stems are used in a combination of contexts which do not form a natural class. Affixal exponents realize unbounded aspect in the indicative (*mi-*), irrealis mood (*be-*), negation (*na-* or *ne-*, not illustrated here), type of nonfinite form (*-e* vs. *-ande* vs. *-an* vs. *-ân*), and subject agreement for finite forms. Within Paradigm Function Morphology, this rather simple position class system can be accounted for using the series of rule

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

Table 1: Rule blocks for Persian synthetic conjugation

blocks outlined in table 1. Remember that in PFM, realization rules are organized in successive blocks. When attempting to realize a given set of morphosyntactic feature, the most specific applicable rule within the block is chosen. (6) are sample rules from block V, written in an attribute-value matrix format.[2]: while (6a) asks that finite verbs with a 2SG subject take the suffix *-i*, the more specific (6b) indicates that the suffix is dropped in the imperative.

(6) a. $\begin{bmatrix}\text{PHON} & X\\ \text{LID} & Y\end{bmatrix}, \sigma : \begin{bmatrix}\text{PER} & 2\\ \text{NB} & sg\end{bmatrix} \longrightarrow \begin{bmatrix}\text{PHON} & X \oplus \text{i}\\ \text{LID} & Y\end{bmatrix}$ (block V)

b. $\begin{bmatrix}\text{PHON} & X\\ \text{LID} & Y\end{bmatrix}, \sigma : \begin{bmatrix}\text{PER} & 2\\ \text{NB} & sg\\ \text{MOOD} & imper\end{bmatrix} \longrightarrow \begin{bmatrix}\text{PHON} & X\\ \text{LID} & Y\end{bmatrix}$ (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 can not easily be formulated in existing description languages for HPSG grammars. Rather than attempting a direct integration, we propose to 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 head values as in (7), where MORSYN groups features that get realized in inflection and LID assigns a specific index to each lexeme (Spencer, 2005; Sag, 2007). 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 feature bundles instead of category structures *à la* (Gazdar et al., 1985). The meta-constraint in (8) then links the two grammars.

[2]Two different conventions are currently used to write PFM rules, defined respectively in (Stump, 2001) and (Ackerman and Stump, 2004). The AVM format we use here is meant to ease the integration with HPSG, although the change is little more than syntactic sugar.

(7) $\textit{head} \rightarrow \begin{bmatrix} \textit{head} & \\ \text{LID} & \textit{lexemic-index} \\ \text{MORSYN} & \textit{morsyn} \end{bmatrix}$

(8) **Morphology-syntax interface (preliminary version)**

A sign of type *word* meeting the description $\begin{bmatrix} \text{PHON} & \boxed{1} & \\ \text{HEAD} & \begin{bmatrix} \text{LID} & \boxed{3} \\ \text{MORSYN} & \boxed{4} \end{bmatrix} \end{bmatrix}$ is well-formed only if the PFM grammar licenses phonology $\boxed{1}$ as a realization of the features $\boxed{4}$ for the lexeme $\boxed{3}$.

2 The passive

The passive in Persian is a typical complex predicate construction, whose properties are parallel to those of copula-predicative complement constructions. The auxiliary *šodan* is clearly the head: all inflectional information, e.g. negation (9), is realized on the auxiliary. The participle-auxiliary sequence is syntactically flexible: adverbs may intervene (10), the auxiliary may be scrambled over the participle (11), and long-distance fronting of the participle is possible (12).

(9) In tâblo foruxte ne-mi-šav-ad.
this painting sold NEG-UNBD-become.S1-3.SG
'This painting is not sold.'

(10) In tâblo foruxte hatman šod.
this painting sold certainly become.S2
'This painting was certainly sold.'

(11) In tâblo šod robude va foruxte.
this painting become.S2 stolen and sold
'It is this painting which was stolen and sold.'

(12) Foruxte fekr mi-kon-am [agar in tâblo _
sold thought UNBD-do.S1-1SG if this painting
be-šav-ad, mi-tavân-im bâ pul-aš yek
SBJV-become.S1-3SG UNBD-can.S1-1PL with money-PAF.3SG a
mâšin be-xar-im].
car SBJV-buy.S1-1PL
'I think that if this painting is sold, we can buy a car with the money.'

To account for this we rely on an argument composition analysis in the spirit of (Hinrichs and Nakazawa, 1994) and subsequent work. Specifically we propose the lexical entry in (13) for the auxiliary lexeme *šodan*, giving rise to analyses such as that in Figure 1. Under our analysis there is no passive participle, and subject demotion is effected directly in the auxiliary's entry. This is appropriate

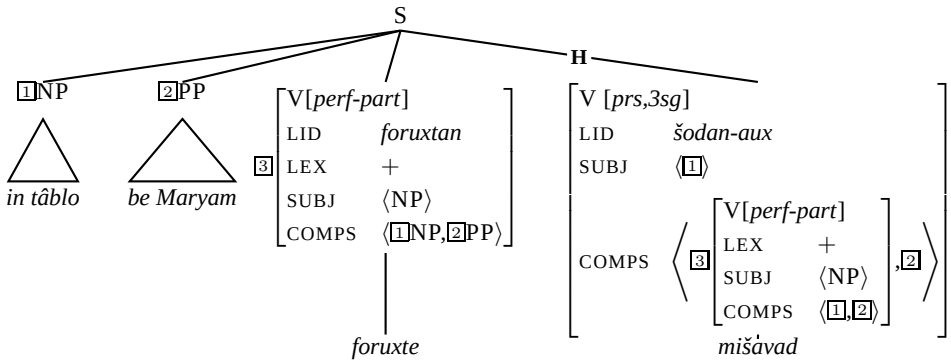

Figure 1: Analysis of a passive sentence

because (i) perfect participles are always active except in the periphrastic passive constructions—participial clauses with transitive head verbs take direct objects (14), and (ii) for semantic reasons there is no hope of using the same lexical entry for the auxiliary *šodan* and the full verb *šodan* (contrary to what happens in languages where the passive auxiliary coincides with the copula). Moreover, we assume a flat structure, wich allows for an easy account of the free reordering of the participle, auxiliary and valents. The specification [VC −] on the participle inhibits the formation of a verbal complex—see below for a contrasting analysis of perfect periphrases.

(13)
$$\begin{bmatrix} \text{HEAD} & \left[\text{LID} \quad \textit{šodan-aux}\right] \\ \text{CONT} & \boxed{2} \\ \text{ARG-ST} & \left\langle \boxed{1}, \begin{bmatrix} \text{FORM} & \textit{part} \\ \text{PERFECT} & + \\ \text{POL} & + \\ \text{CONT} & \boxed{2} \\ \text{ARG-ST} & \langle \text{NP}, \boxed{1} \rangle \oplus \boxed{L} \\ \text{LEX} & + \\ \text{VC} & - \end{bmatrix} \right\rangle \oplus \boxed{L} \end{bmatrix}$$

(14) Maryam tâblo=râ xarid-e va be Omid dâd.
Maryam painting=DDO buy.S2-PRP and to Omid give.S2
'Having bought the painting, Maryam gave it to Omid.'

Notice that under our analysis voice is not an inflectional category in Persian: the active-passive opposition is dealt with entirely within syntax.

3 Two sets of forms based on *budan*

There are five different subparadigms based on *budan*, illustrated here in (15). These contrast in two independent ways.

(15) a. Complex present: *xaride=i*
b. Complex bounded past: *xaride bud-i*
c. Complex unbounded past: *mixaride=i*
d. Complex subjunctive: *xaride bâš-i*
e. Compex perfect: *xaride bude=i*

3.1 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 for 3SG (there is also a nonclitic 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 (here *=i*).

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 (16), as is participle fronting (17). Second, the distribution of the unbounded aspect marker *mi-* is otherwise unexplainable: it is the full construction, not the participle, that is unbounded. Finally, colloquial Persian allows a form of vowel reduction in the 3SG that is peculiar to these forms (18a): comparable contructions where the clitic auxiliary combines with an adjective do not give rise to the same pattern (18b).[3]

(16) *Rafte hatman=ast.
left certainly=be.S1.3SG
'(S)he has certainly left.'

(17) *Ne-mi-rafte sâlhâ Maryam be madrase=ast.
NEG-UNBD-gone years Maryam to school=be.S1.3SG
'For years, Maryam didn't go to school'

(18) a. mord'e=ast → mord'e:
died=be.S1.3SG
'(S)he has died.'

[3]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. However the existence of sublexical coordination in numerous languages calls into question whether this is a strong argument against a morphological analysis. We leave this issue for future research.

b. mord'e=ast → mord'ast
corpse=be.S1.3SG
'It is a corpse.'

Compare now the situation of forms that are based on a nonclitic auxiliary. The participle-auxiliary combination is more constrained than it is in the passive; in particular, neither adverbs (19) nor pronominal affixes (20) can occur between the two verb forms, and negation must be realized on the participle (21). In addition, scrambling is excluded (22). However, the combination is not lexical, since the participle can be extracted (23).

(19) * Maryam dide hatman bud-aš
Maryam seen certainly be.S2-PAF.3SG

(20) a. Maryam dide budaš.
Maryam seen be.S2-PAF.3SG
'Maryam had seen him.'

b. * Maryam dide-aš bud.
Maryam seen-PAF.3SG be.S2

(21) Maryam Omid-râ na-dide bud.
Maryam Omid-DDO NEG-seen be.S2
'Maryam hadn't seen Omid.'

(22) * Maryam Omid-râ bud dide.
Maryam Omid-DDO be.S2 seen

(23) Foruxte fekr ne-mi-kon-am [_ bâš-ad in
sold thought NEG-UNBD-do.S1-1SG be.SBJV-3SG this
tâblo=râ].
painting=DDO
'I don't think that s/he has sold this painting.'

3.2 Morphosyntactic import

The use of a form based on *budan* may realize two distinct morphosyntactic features. The complex bounded past (24) and complex subjunctive (25) 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 bounded past is used when the speaker has direct evidence for what she is asserting, the complex bounded past is used in contexts where the evidence is only indirect, as in (26).

(24) Qabl az inke Omid be-res-ad, Maryam birun rafte bud.
before from that Omid SBJV-arrive.S1-3SG Maryam out gone be.S2
'Maryam had left (before Omid arrived).'

(25) Fekr mi-kon-am Maryam mariz bude baš-ad.
thought UNBD-do.S1-1SG Maryam sick been be.SBJV-3SG
'I think Maryam has been sick.'

(26) (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-built=be.S1.3SG
'According to Omid, Maryam was 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 (27a) or as a bounded past with indirect evidentiality (27b). Finally, the complex perfect expresses both perfectivity and indirect evidentiality: it is the indirect evidential equivalent of the complex bounded past (28). Note that this corresponds transparently to the fact the the complex perfect includes two realizations of the copula.

(27) a. Maryam tâze reside=ast.
Maryam new arrived=be.S1.3SG
'Maryam has just arrived.'

b. (Banâ bar gofte-ye Omid) Maryam in xâne-râ dar sâl-e 1950
According to-EZ Omid) Maryam this house-DDO in year-EZ 1950
xaride=ast.
bought=be.S1.3SG
'According to Omid, Maryam bought this house in 1950.'

(28) (Az qarâr), qabl az inke Omid be-res-ad, Maryam birun
apparently before from that Omid SBJV-arrive.S1-3SG, Maryam out
rafte bude=ast.
gone been=be.S1.3SG
'Apparently, Maryam had left before Omid arrived.'

As can be seen in Table 3.2, if the present perfect is ignored, morphosyntactic properties align nicely with morphologized vs. syntactic combinations: the morphologized forms are used for indirect evidentiality, as stated by rules (29); while the truly periphrastic forms are used to express the perfect. The fact that the present perfect is unexpectedly synthetic calls for an 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. Specifically one should assume a rule of referral along the lines of (30). The rule states that any present perfect form of a lexeme Y will be identical to the indirect bounded past form of Y with the same specifications for all features except tense, aspect and evidentiality (here, the relevant remaining features are person, number and polarity).[4]

[4]This is a portmanteau rule of referral covering blocks I to V, thus bypassing completely synthetic exponence. $\sigma \setminus \tau$ is the description that is identical to σ except where the features mentioned in

	PRESENT	PAST		SBJV
		DIR. EV.	IND. EV.	
BD	***	bounded past	complex present	simple sbjv
UNBD	simple present	unbd past	cpl. unbd. past	
PRF	complex present	complex bnd. past	complex perfect	complex sbjv

Table 2: Morphosyntactic features expressed by Persian subparadigms

(29) a. $\begin{bmatrix}\text{PHON} & X\\ \text{LID} & Y\end{bmatrix}, \sigma : \begin{bmatrix}\text{EVID} & \textit{indir}\end{bmatrix} \longrightarrow \begin{bmatrix}\text{PHON} & X \oplus \mathsf{e}\\ \text{LID} & Y\end{bmatrix}$ (block IV)

b. $\begin{bmatrix}\text{PHON} & X\\ \text{LID} & Y\end{bmatrix}, \sigma : \begin{bmatrix}\text{EVID} & \textit{indir}\\ \text{PER} & \textit{3}\\ \text{NB} & \textit{sg}\end{bmatrix} \longrightarrow \begin{bmatrix}\text{PHON} & X \oplus \mathsf{ast}\\ \text{LID} & Y\end{bmatrix}$ (block V)

(30) $\begin{bmatrix}\text{PHON} & X\\ \text{LID} & Y\end{bmatrix}, \sigma : \begin{bmatrix}\text{TNS} & \textit{prst}\\ \text{PRF} & +\end{bmatrix} \longrightarrow$

$$\begin{bmatrix}\text{PHON} & \mathsf{refer}\left(\begin{bmatrix}\text{PHON} & X\\ \text{LID} & Y\end{bmatrix}, \sigma \setminus \begin{bmatrix}\text{TNS} & \textit{pst}\\ \text{ASP} & \textit{bnd}\\ \text{PRF} & -\\ \text{EVID} & \textit{ind}\end{bmatrix}, \text{I-V}\right)\\ \text{LID} & Y\end{bmatrix}$$ (blocks I-V)

4 Analyzing the perfect periphrases

We construct the analysis of perfect periphrases in two steps. First we present a syntactic analysis of perfect forms based on argument composition, and show what is unsatisfactory with such an approach. Next we present a way of arriving at the same syntactic analysis by inflectional means. Finally we discuss alternatives and potential problems.

τ differ from those in σ. The function refer takes as arguments an indexed phonological form, a morphosyntactic specification and a rule block sequence, and outputs the result of applying to this indexed phonological form and this morphosyntactic specification the restriction of the PFM grammar to these rule blocks. The motivation for deriving the present perfect from the indirect bounded past rather than the other way around is the economy of paradigms: this allows us to state the rules of exponence realizing suffixes *-e* and *-ast* in a natural way, as applying to all and only evidential forms. Notice that the orientation of the rule of referral might not correspond to the directionality of the diachronic morphologization process.

4.1 A failed analysis based on argument composition

As a first step, we present an analysis that is a variation of the analysis presented above for the passive. (31) is a candidate entry for the present form of the auxiliary *bud*. This states that the auxiliary is a past perfect form which takes a perfect participle complement and inherits the participle's arguments. Because the past participle is marked as [VC +], the auxiliary and participle form a verbal complex, as indicated in figure 2 and thus can not be seperated by elements that are not allowed to occur inside a verbal complex. Rigid word order is a consequence of the LP rule in (32). In addition, since the participle is an argument of the auxiliary, this analysis will allow for the extraction of the participle within any HPSG approach to extraction.

(31)
$$\begin{bmatrix} \text{PHON} & \text{bud} \\ \text{HEAD} & \begin{bmatrix} \text{LID} & \textit{budan-aux} \\ \text{MORSYN} & \begin{bmatrix} \textit{tns} & \textit{pst} \\ \text{PRF} & + \\ \text{AGR} & \begin{bmatrix} \text{PER} & 3 \\ \text{NB} & sg \end{bmatrix} \\ \text{POL} & + \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & \left\langle \boxed{1}, \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{FORM} & \textit{part} \\ \text{PRF} & + \\ \text{POL} & + \end{bmatrix} \\ \text{LEX} & + \\ \text{VC} & + \\ \text{ARG-ST} & \langle \boxed{1} \rangle \oplus \boxed{L} \end{bmatrix} \right\rangle \oplus \boxed{L} \end{bmatrix}$$

(32)
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{FORM} & \textit{part} \\ \text{PRF} & + \\ \text{VC} & + \end{bmatrix} \end{bmatrix} \prec [\ ]$$

While this analysis is appropriate as far as syntax is concerned, its integration with the analysis of synthetic conjugation is problematic. First, the perfect auxiliary must be stipulated to be defective for all nonperfect forms, and to have perfect forms that are homonymous to the nonperfect forms of the ordinary copula; thus the purported perfect auxiliary is inflectionally deeply abnormal. Second, we need to derive the fact that there is no present form of the perfect auxiliary (remember that the present perfect is a morphologized form). There are two ways this could be done. We could further stipulate that the perfect auxiliary is defective for

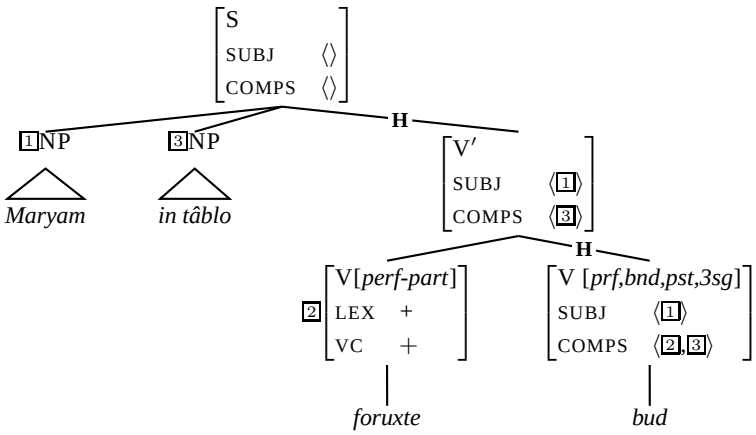

Figure 2: The syntactic structure of perfect periphrases

the present, despite the fact that the ordinary copula it derives from has perfectly good present forms (in fact, two sets of such forms: clitic and nonclitic ones). Or we could assume that some form of competition between morphology and syntax is taking place (Poser, 1992)—but the postulation of such competitions is notoriously difficult to state precisely, and quite alien to the design properties of HPSG. Finally, we need to find a way of stating that the passive auxiliary can not take the perfect auxiliary as its complement: while (33a) is well-formed, (33b) is not.

(33) a. In tâblo foruxte šode bud.
this painting sold become be.PST.3SG
'This painting had been sold.'

b. * In tâblo foruxte bude šod.
this painting sold be become.PST.3SG

While these problems can definitely be circumvented by specifying an appropriately complex inflectional paradigm for the perfect auxiliary, it is striking that many conterintuitive stipulations are needed just because it is not possible to state that the periphrastic perfect is part of the inflectional paradigm of the main verb. The next subsection attempts to modify the framework in a way that allows for the formulation of such an analysis.

4.2 An alternative solution: exponence as valence

As the last subsection stressed, what we need is a way to treat perfect forms as part of the inflectional paradigm (Ackerman and Stump, 2004), 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 *xaridan* meets the description in (34), which is exactly like (31) except for the fact that it is an instance of the lexeme *xaridan*.

(34)
$$\begin{bmatrix}
\text{PHON} & \textsf{bud} \\
\text{HEAD} & \begin{bmatrix} \text{LID} & \textit{xaridan} \\ \text{MORSYN} & \begin{bmatrix} \textit{tns} & \textit{pst} \\ \text{PRF} & + \\ \text{AGR} & \boxed{2} \\ \text{POL} & + \end{bmatrix} \end{bmatrix} \\
\text{ARG-ST} & \left\langle \boxed{1}, \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{FORM} & \textit{part} \\ \text{PRF} & + \\ \text{POL} & + \end{bmatrix} \\ \text{LEX} & + \\ \text{VC} & + \\ \text{ARG-ST} & \langle \boxed{1} \rangle \oplus \boxed{L} \end{bmatrix} \right\rangle \oplus \boxed{L}
\end{bmatrix}$$

The challenge now is to derive (34) 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 valence lists be added to the picture: realization rules now relate triplets of a phonological representation, a lexemic index, and an argument structure specification. The meta-constraint in (8) is updated as in (35), so that argument structure is examined at the morphology-syntax interface.[5]

(35) **Morphology-syntax interface (preliminary 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 arguments $\boxed{2}$ as a realization of the features $\boxed{4}$ for the lexeme $\boxed{3}$.

The rule licensing (34) is given in (36). To realize a feature structure σ verifying [PRF +], one should refer the phonology to that of the corresponding bounded

[5]The formulation of this constraint presupposes that the HPSG grammar says nothing about individual lexical entries, and that most of the usual HPSG theory of the lexicon is recast as part of the morphological component.

positive nonperfect form of *budan*, and add to the argument list a requirement for a form of Y realizing the same feature set except for the fact that it is a participle.

(36)
$$\begin{bmatrix} \textsc{phon} & X \\ \textsc{lid} & Y \\ \textsc{val} & Z \end{bmatrix}, \sigma : \begin{bmatrix} \textsc{prf} & + \end{bmatrix} \longrightarrow$$
$$\begin{bmatrix} \textsc{phon} & \textsf{refer}\left(\begin{bmatrix} \textsc{phon} & X \\ \textsc{lid} & \textit{budan} \\ \textsc{val} & Z \end{bmatrix}, \sigma \backslash \begin{bmatrix} \textsc{prf} & - \\ \textsc{asp} & \textit{bnd} \\ \textsc{pol} & + \end{bmatrix}, \text{I-V} \right) \\ \textsc{lid} & Y \\ \textsc{val} & Z \oplus \left\langle \begin{bmatrix} \textsc{lex} & + \\ \textsc{vc} & + \\ \textsc{head} & \begin{bmatrix} \textsc{lid} & Y \\ \textsc{morsyn} & \sigma \backslash \begin{bmatrix} \textsc{form} & \textit{part} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

The proposed analysis makes the following correct predictions. First, negation is handled correctly: the phonology of the head word is constrained to be that of a positive form of *budan*, whereas the participle shares its polarity value with that of the head word. Thus the head will never carry a negation prefix, but its negative polarity value will be realized as a prefix on the participle it selects. Second, the complex perfect is predicted to exist without stipulation: because evidentiality is morphologized and available for all past forms, rule (36) will generate an indirect past perfect with the phonology of an indirect bounded past form of *budan*. Figure 3 illustrates the relevant analysis. Third, the analysis correctly predicts that (33a), and not (33b), is grammatical. This is because the passive auxiliary, as a lexeme, can be put in the perfect; whereas there is no sense in which one can put the perfect auxiliary in the passive, because there is no such thing as a perfect auxiliary lexeme. The analysis of (33a) is shown in Figure 4.

Finally we account straightforwardly for the nonexistence of a periphrastic present perfect. Since (36) is an inflectional realization rule, it interacts with the rule of referral in (30) under the logic of rule specificity: thus the existence of (30) overrides the application of (36). In this sense the current analysis of the periphrastic perfect is syntactically reductionist: periphrasis is reduced to valence; no phrasal constructions or lexical entries are needed. We assume a notion of rule competition, but this competition is segregated to the inflectional component, where it is arguably needed for independent reasons. Thus no competition between morphology and syntax (e.g. Poser, 1992; Bresnan, 2001) needs to be orcherstrated.

4.3 Discussion

The analysis of the Persian perfect outlined above attempts to capture the traditional intuition of periphrastic inflection. While there are many ways one might

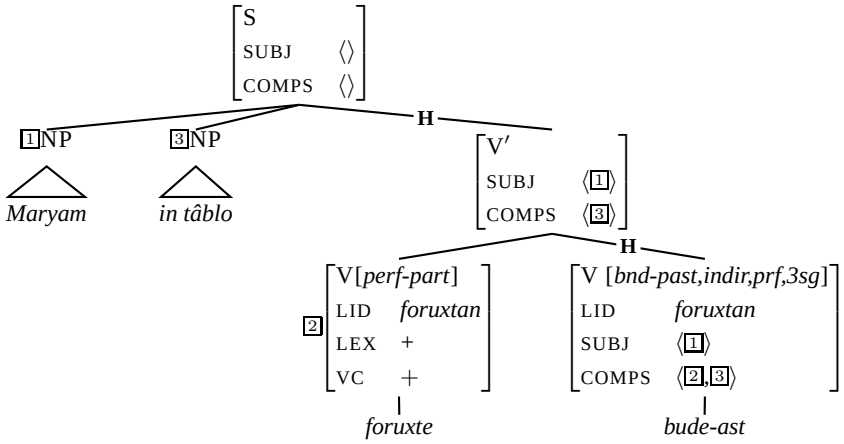

Figure 3: Analysis of a sentence in the complex pefect: ‘Reportedly, Maryam had sold this painting.’

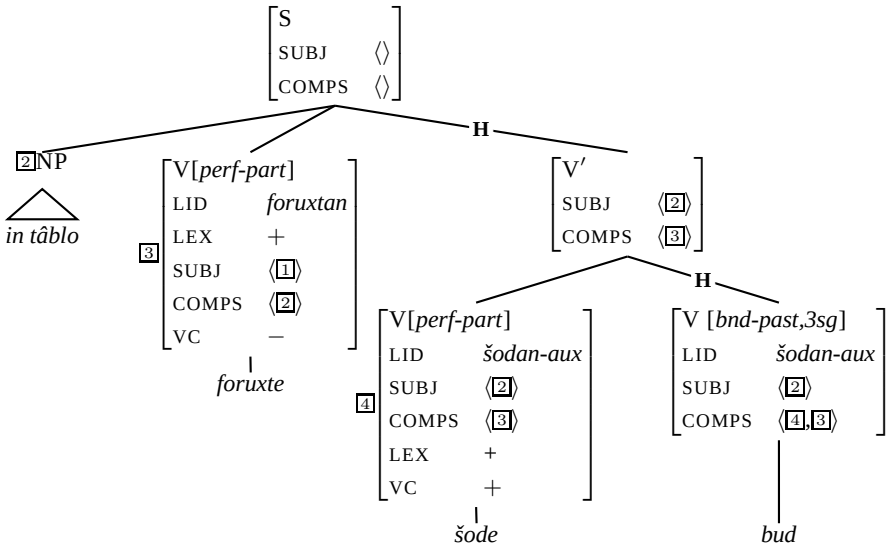

Figure 4: Perfect-passive interaction: analysis for (33a)

attempt to reach this goal in the context of HPSG (see in particular Ackerman and Webelhuth, 1998), the specific design goal here has been to devise an analysis that meets as much as possible both the analytical habits of HPSG syntax and of realizational morphology. Thus as far as clausal syntax is concerned, our analysis is undistinguishable from an argument composition analysis, and we have attempted to account for all relevant syntactic features of the construction. On the other hand, the lexical analysis is as close as possible to standard realizational morphology; in particular it relies heavily on the architecture of rule blocks and rule competition to generate the correct lexical representations.

While we fully assume this research strategy, alternatives are readily conceivable that meet different design goals but produce very similar analyses. For instance, turning the analysis into a standard HPSG analysis, with item-and-process morphology encoded via lexical rules, is easy: one just needs to recast rules such as (30) and (36) as lexical rules, and modify the morphosyntactic descriptions (using quite a bit of negation and disjunction and/or fine-tuning the type system) so as to make them mutually exclusive. The resulting system is more conservative from an HPSG perspective, although one may doubt that it is more perspicuous.

A different issue raised by the current analysis is its interaction with the analysis of coordination. Coordination of participles is possible in the perfect in Persian, just as it is in the passive (37). This can not be treated as a simple instance of constituent coordination under our analysis: because we assume that the auxiliary is really an inflected form of the main verb, there is no single lexeme of which *bud* is the realization in (37). While this is definitely a problem, it is a familiar one, reminiscent of issues pertaining to coordinations of unlikes. We see two potential solutions. First, we could assume an ellipsis-based analysis of (37) along the lines of analyses proposed by (Yatabe, 2001; Crysmann, 2003; Beavers and Sag, 2004). Second, we could assume a richer ontology of LID values where a neutralized value common to both participles is assigned to the coordinate phrase in (37), extending work in the tradition of (Daniels, 2002; Levy and Pollard, 2002; Sag, 2003). This neutralized value could then serve as an appropriate input for rule (36).[6] Whether these strategies prove fruitful will have to wait for future research, and in particular for a detailed empirical study of coordination in Persian.

(37) Maryam tâblo-râ pasandide va xaride bud.
Maryam painting-DDO liked and bought be.PST
'Maryam had liked and bought the painting.'

5 The future

For the periphrastic future, a number of different analytic options are available. As in the case of the periphrastic perfect, the verb sequence can not be interrupted, and

[6]Notice that the postulation of neutralized LID values is needed anyway to allow for constituent coordination under the assumptions of (Sag, 2007). Thus the issue raised by our analysis is an issue that needs to be addressed anyway.

occurs in a rigid order.

(38) a. Maryam Omid=râ xâh-ad did.
Maryam Omid=DDO want.S1-3.SG see.S2
'Maryam will see Omid.'

b. *Maryam xâh-ad Omid=râ did.
Maryam want.S1-3.SG Omid=DDO see.S2

c. *Maryam Omid-râ did xâh-ad.
Maryam Omid-DDO see.S2 want.S1-3.SG

The periphrastic future does not enter into paradigmatic relations with syncretic inflection. Thus it could be accounted for entirely within syntax. On the other hand, syntactic rules do not manipulate portions of the periphrastic construction—notably, the nonauxiliary part of the future can not be fronted. Thus nothing precludes either a purely morphological analysis.

There is however one argument favouring a purely morphological analysis, although it is not a very strong one. The future auxiliary looks like a present tense form of *xâstan* 'want', except that it does not carry the unbounded auxiliary normally found in the present. If we were to treat the future construction as phrasal, we would thus need to set up the grammar so that the morphology output supplementary forms, the distribution of which we would then need to constrain drastically within syntax.[7] We thus opt for a purely morphological analysis.[8] We propose to use the rule in (39), which is a double portmanteau rule of referral. To find the

[7]The nonfinite form appears to be a bare past stem. Words homophonous to a bare past stem are used in two other contexts: in the bounded past with a 3SG subject, where the exponent of agreement is null; and in the impersonal complement of some modal verbs such as *bâyastan* 'must, be necessary' (1).

(1) a. Maryam (hatman) bây-ad be madrasa be-rav-ad.
Maryam certainly must.S1-3SG to school IRR-go.S1-3SG
'Maryam definitely has to go to school.'

b. (Hatman) bây-ad be madrase raft.
certainly must.S1-3SG to school go.S2
'It is definitely necessary to go to school.'

[8]One could argue from the fact that object clitics can be realized either on the auxiliary (i) or on the nonfinite form (ii) that they should be treated as two distinct syntactic atoms; but since we treat object clitics as affixes anyway, the question is moot. In any case, the analysis in (39) can readily be extended to account for (i), but an account of (ii) will need to rely on more extensive revisions.

(i) Maryam xâh-ad did-aš
Maryam want.S1-3.SG see.S2-PAF.3.SG
'Maryam will see her/him.'

(ii) Maryam xâh-ad-aš did
Maryam want.S1-3.SG-PAF.3.SG see.S2
'Maryam will see her/him.'

phonology of a future form, one needs to concatenate the output of block IV on the form *xâh* with a bare past stem of the lexeme being realized.

(39)
$$\begin{bmatrix} \textsc{phon} & X \\ \textsc{lid} & Y \\ \textsc{arg-st} & Z \end{bmatrix}, \sigma : \left[\textsc{tns} \quad \textit{fut}\right] \longrightarrow$$
$$\begin{bmatrix} \textsc{phon} & \textsf{refer}\left(\begin{bmatrix} \textsc{phon} & \textsf{xâh} \\ \textsc{lid} & Y \\ \textsc{arg-st} & Z \end{bmatrix}, \sigma, \textsc{v}\right) \oplus \\ & \textsf{refer}\left(\begin{bmatrix} \textsc{phon} & X \\ \textsc{lid} & Y \\ \textsc{arg-st} & Z \end{bmatrix}, \sigma \backslash \left[\textsc{tns} \quad \textit{pst}\right], \textsc{i}\right) \\ \textsc{lid} & Y \\ \textsc{arg-st} & Z \end{bmatrix}$$

6 The progressive

All unbounded forms may give rise to a progressive interpretation, but that interpretation can also be forced by using the periphrastic construction illustrated in (4). Unlike the ones we discussed so far, this construction 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.[9] The nonauxiliary 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, but complementizers are optional for finite complements (40). Complements normally occur between the two verbs; they can scramble to the left of the auxiliary, but this is also possible with clausal complements (41). Finally, object clitic pronouns must be realized on the nonauxiliary verb, and cannot climb to the auxiliary (42).

(40) a. Maryam dâr-ad (*ke) ketâb mi-xân-ad
Maryam have.S1-3SG COMP book UNBD-read.S1-3SG
'Maryam is reading a book.'

b. Maryam mi-xâh-ad (ke) bâ Omid har ruz be sinemâ be-rav-ad
Maryam UNBD-want.S1-3SG COMP with Omid every day to theatre SBJV-go.S1-3SG
'Maryam wants to go to theatre with Omid everyday.'

[9] Persian raising and control constructions normally rely on a finite unsaturated complement clause. Infinitival complements are available only in a very formal register.

(41) a. Maryam in ketâb=râ dâr-ad mi-xân-ad
Maryam this book=DDO have.S1-3SG UNBD-read.S1-3SG
'Maryam is reading this book.'

b. Maryam bâ Omid mi-xâh-ad (ke) har ruz be sinemâ be-rav-ad
Maryam with Omid UNBD-want.S1-3SG COMP every day to theatre SBJV-go.S1-3SG
'Maryam wants to go to theatre with Omid everyday.'

(42) a. Maryam dâr-ad mi-xân-ad=aš
Maryam have.S1-3.SG UNBD-read.S1-3SG-=3SG
'Maryam is reading it.'

b. * Maryam dâr-ad=aš mi-xân-ad
Maryam have.S1-3SG=3SG UNBD-read.S1-3SG

This data can be accounted for by assuming a slightly idiosyncratic lexemic entry for the auxiliary *dâštan*. This entry assumes that *prog* is a subtype of the ASPECT value *unbd* (unbounded). As a result of its lexeme-level specification, this auxiliary is defective for all subparadigms except the present, the unbounded past and the complex unbounded past, in accordance with the facts. The subject of the complement is constrained to be an *nc-pro*, the type of pro-dropped subjects, and coindexed with the auxiliary's subject. The analysis is illustrated in Figure 5.

(43)
$$\left[\begin{array}{ll}
\text{HEAD} & \left[\begin{array}{ll}\text{LID} & \textit{dâštan-aux}\\ \text{MORSYN} & \boxed{1}\left[\text{ASP} \quad \textit{prog}\right]\end{array}\right]\\
\text{CONT} & \boxed{2}\\
\text{ARG-ST} & \left\langle\left[\text{IND} \quad \boxed{3}\right], \left[\begin{array}{ll}\text{MORSYN} & \boxed{1}\\ \text{MARKING} & \textit{none}\\ \text{CONT} & \boxed{2}\\ \text{SUBJ} & \left\langle\left[\begin{array}{ll}\textit{nc-pro}\\ \text{IND} & \boxed{3}\end{array}\right]\right\rangle\\ \text{COMPS} & \langle\rangle\end{array}\right]\right\rangle
\end{array}\right]$$

References

Ackerman, Farrell and Stump, Gregory T. 2004. Paradigms and Periphrastic Expression. In Louisa Sadler and Andrew Spencer (eds.), *Projecting Morphology*, pages 111–157, Stanford: CSLI Publications.

Ackerman, Farrell and Webelhuth, Gert. 1998. *A Theory of Predicates*. Stanford: CSLI Publications.

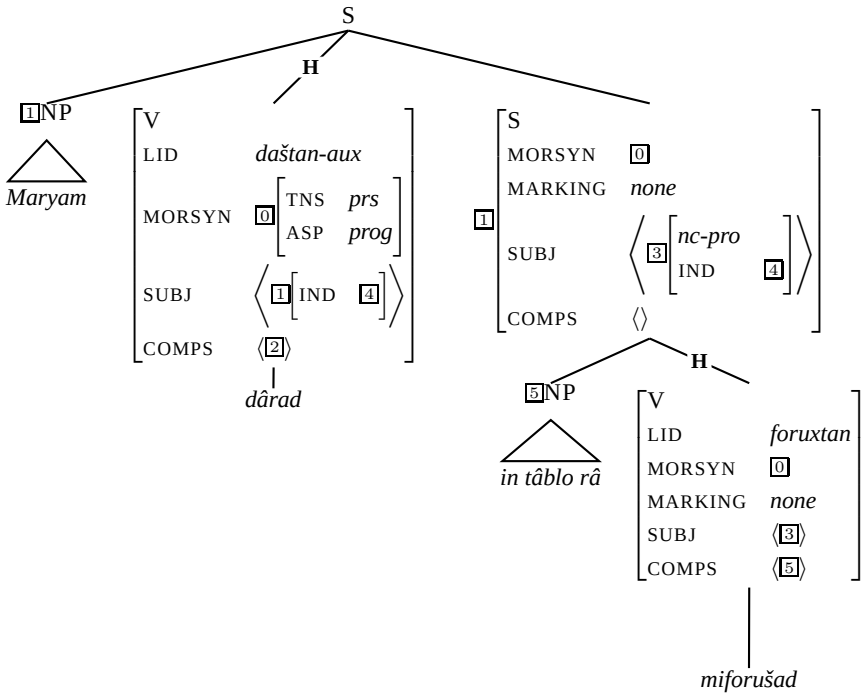

Figure 5: Analysis for (4) ‘Maryam is selling the painting.’

Beavers, John and Sag, Ivan A. 2004. Coordinate Ellipsis and Apparent Non-Constituent Coordination. In *Proceedings of the HPSG04 Conference*, Stanford: CSLI Publications.

Bresnan, Joan. 2001. *Lexical-functional syntax*. Oxford: Basil Blackwell.

Crysmann, Berthold. 2003. An Asymmetric Theory of Peripheral Sharing in HPSG: Conjunction Reduction and Coordination of Unlikes. In *Proceedings of Formal Grammar 2003*.

Daniels, Michael W. 2002. On a Type-Based Analysis of Feature Neutrality and the Coordination of Unlikes. In Frank Van Eynde, Lars Hellan and Dorothee Beermann (eds.), *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pages 137–147, Stanford: CSLI Publications.

Gazdar, Gerald, Klein, Ewan, Pullum, Geoffrey K. and Sag, Ivan A. 1985. *Generalized Phrase Structure Grammar*. Harvard: Harvard University Press.

Hinrichs, Erhard and Nakazawa, Tsuneko. 1994. Linearizing AUXs in German verbal complexes. In John Nerbonne, Klaus Netter and Carl Pollard (eds.), *German in HPSG*, pages 11–37, Stanford: CSLI Publications.

Jahani, Carina. 2000. Expressions of indirectivity in spoken Modern Persian. In Lars Johanson and Bo Utas (eds.), *Evidentials: Turkic, Iranian and Neighbouring Languages*, pages 185–207, Berlin: Mouton de Gruyter.

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

Lazard, Gilbert, Richard, Yann, Hechmati, Rokhsareh and Samvelian, Pollet. 2006. *Grammaire du persan contemporain*. Tehran: IFRI and Farhag-e Moaser.

Levy, Roger and Pollard, Carl. 2002. Coordination and Neutralization in HPSG. In Frank Van Eynde, Lars Hellan and Dorothee Beermann (eds.), *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pages 221–230, Stanford: CSLI Publications.

Poser, William J. 1992. Blocking of phrasal constructions by lexical items. In Ivan A. Sag and Anna Szabolcsi (eds.), *Lexical matters*, pages 111–130, Stanford: CSLI Publications.

Sag, Ivan A. 2003. Coordination and underspecification. In *Proceedings of the 9th HPSG Conference*, pages 267–291, Stanford: CSLI Publications.

Sag, Ivan A. 2007. SIgn-Based Construction Grammar: An Informal Synopsis, ms., Stanford University.

Spencer, Andrew. 2005. Inflecting clitics in Generalized Paradigm Function Morphology. *Lingue e Linguaggio* 4, 179–193.

Stump, Gregory T. 2001. *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

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

Yatabe, Shûichi. 2001. The Syntax and Semantics of Left-Node Raising in Japanese. In Dan Flickinger and Andreas Kathol (eds.), *Proceedings of the 2001 HPSG Conference*, Stanford: CSLI Publications.