

Inflectional periphrasis in Persian

Olivier Bonami Pollet Samvelian

U. Paris-Sorbonne

U. Paris 3

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 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 become.S1-3SG sell.S2
‘Maryam will sell the painting’
- (4) Maryam dâr-ad in tâblo=râ
Maryam have.PRS-3SG this painting=DDO
mi-foruš-ad.
UNBD-sell.S1-3SG
‘Maryam is selling the painting.’

The differing properties of these four 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 four 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.

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

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 system of rule blocks outlined in (6). 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. (7) are sample rules from block V, written using the conventions of citepAckerman04: while (7a) asks that finite verbs with a 2.SG subject take the suffix *-i*, the more specific (7b) indicates that the suffix is dropped in the imperative.

	III	II	I	IV	V
	<i>na-</i>	<i>mi-</i>	stem-selection	<i>-e</i>	<i>-am</i>
	<i>ne-</i>			<i>-ande</i>	<i>-i/∅</i>
(6)	<i>be-</i>			<i>an</i>	<i>-ad/∅</i>
					<i>-im</i>
					<i>-id</i>
					<i>-and</i>

- (7) a. $X_V, \sigma : \{\text{PER } 2, \text{NB } \textit{sg}\} \rightarrow X_i$
b. $X_V, \sigma : \{\text{PER } 2, \text{NB } \textit{sg}, \text{MOOD } \textit{imper}\} \rightarrow X$

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 (8), 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 (i) typed feature structures are used to model morphosyntactic feature bundles instead of category structures *à la* (Gazdar et al., 1985); and (ii) the rules output valence lists in addition to phonological representations (see below section 4 for motivation). The meta-constraint in (9) then links the two grammars.

$$(8) \text{ head} \rightarrow \left[\begin{array}{l} \text{head} \\ \text{LID} \quad \text{lexemic-index} \\ \text{MORSYN} \quad \text{morsyn} \end{array} \right]$$

(9) A sign of type *word* meeting the description

$$\left[\begin{array}{l} \text{PHON} \quad [1] \\ \text{VAL} \quad [2] \\ \text{HEAD} \quad \left[\begin{array}{l} \text{LID} \quad [3] \\ \text{MORSYN} \quad [4] \end{array} \right] \end{array} \right]$$

is well-formed only if the PFM grammar licenses phonology [1] and valence [2] as a realization of the features [4] for the lexeme [3].

2 The passive

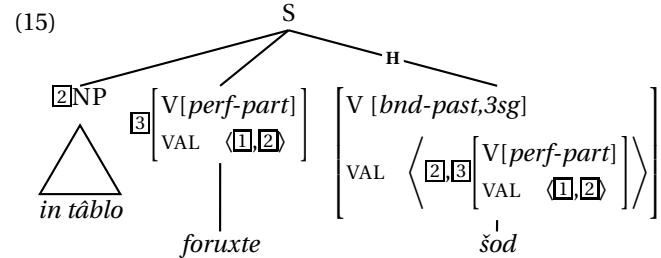
The passive in Persian is a typical complex predicate construction, whose properties are parallel to those of copula-predicative complement constructions. The auxiliary *šodan* is clearly the head: all inflectional information, e.g. negation (10), is realized on the auxiliary. The participle-auxiliary sequence is syntactically flexible: adverbs may intervene (11), the auxiliary may be scrambled over the participle (12), and long-distance fronting of the participle is possible (13).

- (10) In *tâblo* *foruxte* *ne-mi-šav*-ad.
 this painting sold NEG-UNBD-become.S1-3.SG
 ‘This painting is not sold.’
- (11) In *tâblo* *foruxte* *hatman* *šod*.
 this painting sold certainly become.S2
 ‘This painting was certainly sold.’
- (12) In *tâblo* *šod* *robude va* *foruxte*.
 this painting become.S2 stolen and sold
 ‘It is this painting which was stolen and sold’
- (13) *Foruxte* *fekr* *mi-kon-am* [*agar in tâblo*
 sold thought UNBD-do.S1-1SG if this painting
 __ *be-šav-ad* (...)]
 SBJV-become.S1-3SG
 ‘I think that if this painting is sold (...).’

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

(14) for the auxiliary lexeme *šodan*, giving rise to analyses such as that in (15). Note that under our analysis there is no passive participle, and subject demotion is effected directly in the auxiliary’s entry. This is appropriate because (i) perfect participles are always active except in the periphrastic passive constructions (e.g. participial clauses with transitive head verbs take direct objects), 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). Finally, note that under our analysis voice is not an inflectional category in Persian: the active-passive opposition is dealt with entirely within syntax.

$$(14) \left[\begin{array}{l} \text{HEAD} \quad \left[\begin{array}{l} \text{LID} \quad \text{šodan-aux} \end{array} \right] \\ \text{CONT} \quad [1] \\ \text{VAL} \quad [L] \oplus \left\langle \begin{array}{l} \text{FORM} \quad \text{part} \\ \text{PERFECT} \quad + \\ \text{POL} \quad + \\ \text{CONT} \quad [1] \\ \text{VAL} \quad \langle \text{NP} \rangle \oplus [L] \\ \text{LEX} \quad + \end{array} \right\rangle \end{array} \right]$$



3 Two sets of forms based on *budan*

There are five different subparadigms based on *budan*, illustrated here in (16). These contrast in two independent ways.

- (16) 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. Complex 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 (17), as is participle fronting (18). 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 (19a): comparable constructions where the clitic auxiliary combines with an adjective do not give rise to the same pattern (19b).²

- (17) *Rafte hatman=ast.
left certainly=be.s1.3SG
'(S)he has certainly left.'
- (18) *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'
- (19) a. raft'e=ast → raft'e:
left=be.s1.3SG
'(S)he left.'
- b. bard'e=ast → bard'ast
slave=be.s1.3SG
'(S)he is a slave.'

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 (20) nor object clitics (21) can occur between the two verb forms, and negation must be realized on the participle (22). In addition, scrambling is excluded (23). However, the combination is not lexical, since the participle can be extracted (24).

- (20) *Maryam dide hatman bud=aš
Maryam seen certainly be.s2=3SG
- (21) a. Maryam dide bud=aš.
Maryam seen be.s2=3SG
'Maryam had seen him.'
- b. *Maryam dide-aš bud.
Maryam seen=3SG be.s2
- (22) Maryam Omid-râ na-dide bud.
Maryam Omid-DDO NEG-seen be.s2
'Maryam hadn't seen Omid.'
- (23) *Maryam Omid-râ bud dide.
Maryam Omid-DDO be.s2 seen
- (24) Foruxte fekr ne-mi-kon-am [_
sold thought NEG-UNBD-do.s1-1SG
bâš-ad in tâblo=râ].
be.SBJV-3SG this painting=DDO
'I don't think that s/he has sold this painting.'

²The 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.

3.2 Semantic import

The use of a form based on *budan* may have two types of semantic import. The complex bounded past (25) and complex subjunctive (26) 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 (27).

- (25) Qabl az inke Omid be-res-ad, Maryam
before from that Omid SBJV-arrive.s1-3SG Maryam
birun rafte bud.
out gone be.s2
'Maryam had left (before Omid arrived).'
- (26) Fekr mi-kon-am Maryam mariz bude
thought UNBD-do.s1-1SG Maryam sick been
bašad.
be.SBJV-3SG
'I think Maryam has been sick.'
- (27) (Banâ bar gofte-ye Omid) Maryam dar sâl-e 1950
According to-EZ Omid Maryam in year-EZ 1950
in xâne-râ mi-sâxte=ast.
this house-DDO UNBD-built=be.s1.3SG
'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 (28a) or as a bounded past with indirect evidentiality (28b). Finally, the complex perfect expresses both perfectivity and indirect evidentiality: it is the indirect evidential equivalent of the complex bounded past (29). Note that this corresponds transparently to the fact the the complex perfect includes two realizations of the copula.

- (28) a. Maryam tâze reside=ast.
Maryam new arrived=be.s1.3SG
'Maryam has just arrived.'
- b. (Banâ bar gofte-ye Omid) Maryam in
According to-EZ Omid) Maryam this
xâne-râ dar sâl-e 1950 xaride=ast.
house-DDO in year-EZ 1950 bought=be.s1.3SG
'According to Omid, Maryam bought this house in 1950.'
- (29) (Az qarâr), qabl az inke Omid
apparently before from that Omid
be-res-ad, Maryam birun rafte
SBJV-arrive.s1-3SG, Maryam out gone
bude=ast.
been=be.s1.3SG
'Apparently, Maryam had left before Omid arrived.'

As can be seen in (30), 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 (31); 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. 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 and Stump, 2004).

(30)

	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

(31) $X_V, \sigma : \{EVID\ indir\} \rightarrow Xe$ (block IV)

$X_V, \sigma : \{EVID\ indir, 3sg\} \rightarrow Xast$ (block V)

4 Valence as exponence

As the last section stressed, 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 *xaridan* meets the description in (32). Notice the discrepancy between the lexemic index and the phonology. Because of the VC + specification, (32) gives rise to a verbal complex, which contrasts with our analysis if the passive (15) and accounts for the tighter solidarity of the verbal sequence. In addition, it will allow for the extraction of the participle within any HPSG approach to extraction.

(32)

PHON	bud		
HEAD	MORSYN	LID	<i>xaridan</i>
		TNS	<i>past</i>
VAL	$\left\langle \begin{array}{l} NP, NP \\ \end{array} \right\rangle$	PRF	+
		AGR	<i>3sg</i>
		POL	+
		LEX	+
		VC	+
		LID	<i>xaridan</i>
		MORSYN	$\left[\begin{array}{l} FORM\ part \\ PRF\ + \\ POL\ + \end{array} \right]$

The challenge now is to derive (32) 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. The rule licensing (32) is given in (33). To realize a feature structure σ verifying [PRF +], one should refer the phonology to that of the corresponding bounded positive nonperfect form of *budan*, and add to the valence list a requirement for a form of Y realizing the same feature set except for the fact that it is a participle.³

(33)

PHON	X		
LID	Y	$\sigma : [PRF\ +]$	\rightarrow
VAL	Z		
PHON	refer	$\left(\left[\begin{array}{l} PHON\ X \\ LID\ budan \\ VAL\ Z \end{array} \right], \sigma \setminus \left[\begin{array}{l} PRF\ - \\ ASP\ bnd \\ POL\ + \end{array} \right], I-V \right)$	
LID	Y		
VAL	$Z \oplus$	$\left\langle \begin{array}{l} LEX\ + \\ VC\ + \\ HEAD\ \left[\begin{array}{l} LID\ Y \\ MORSYN\ \sigma \setminus [FORM\ part] \end{array} \right] \end{array} \right\rangle$	

This approach to periphrastic perfect forms has two definite advantages. First, it is syntactically reductionist: periphrasis is reduced to valence; no phrasal constructions or lexical entries are needed; no competition between morphology and syntax (e.g. Poser, 1992; Bresnan, 2001) needs to be orchestrated. Second, since (33) 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 in (34), which overrides the application of (33) because of specificity.

(34)

PHON	X		
LID	Y	$\sigma : \left[\begin{array}{l} TNS\ prst \\ PRF\ + \end{array} \right]$	\rightarrow
VAL	Z		
PHON	refer	$\left(\left[\begin{array}{l} PHON\ X \\ LID\ Y \\ VAL\ Z \end{array} \right], \sigma \setminus \left[\begin{array}{l} ASP\ bnd \\ PRF\ - \\ EVID\ ind \end{array} \right], I-V \right)$	
LID	Y		
VAL	Z		

5 The future

The future shares many of the properties of the periphrastic perfect forms, two notable exceptions being the morphological makeup of the auxiliary and nonfinite form, and the order

³This 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 τ differ from those in σ . As the reader can check, the meta-constraint in (9) makes sure that the valence requirement added by the realization rule will indeed constrain the valence of the word within the HPSG grammar.

between them (see the introduction). It can be accounted for using a rule that is quite similar to (33), and LP rules sensitive to verb form to account for the differing order. For lack of space we can not provide more details in this abstract.

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.⁴ 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 (35). Complements normally occur between the two verbs; they can scramble to the left of the auxiliary, but this is also possible with clausal complements (36). Finally, object clitic pronouns must be realized on the nonauxiliary verb, and cannot climb to the auxiliary (37).

- (35) a. Maryam dâr-ad (*ke) ketâb
Maryam have.S1-3SG COMP book
mi-xân-ad
UNBD-read.S1-3SG
'Maryam is reading a book.'
- b. Maryam mi-xân-ad (ke) bâ Omid
Maryam UNBD-want.S1-3SG COMP with Omid
har ruz be sinemâ be-rav-ad
every day to theatre SBJV-go.S1-3SG
'Maryam wants to go to theatre with Omid every day.'
- (36) a. Maryam in ketâb=râ dâr-ad
Maryam this book=DDO have.S1-3SG
mi-xân-ad
UNBD-read.S1-3SG
'Maryam is reading this book.'
- b. Maryam bâ Omid mi-xân-ad (ke)
Maryam with Omid UNBD-want.S1-3SG COMP
har ruz be sinemâ be-rav-ad
every day to theatre SBJV-go.S1-3SG
'Maryam wants to go to theatre with Omid every day.'
- (37) 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 *dâstan*. This entry

⁴Persian raising and control constructions normally rely on a finite unsaturated complement clause. Infinitival complements are available only in a very formal register.

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.

$$(38) \left[\begin{array}{l} \text{HEAD} \left[\begin{array}{l} \text{LID} \quad dâstan\text{-aux} \\ \text{MORSYN} \quad \boxed{1} [\text{ASP} \quad prog] \end{array} \right] \\ \text{CONT} \quad \boxed{2} \\ \text{VAL} \quad \left\langle \boxed{3}, \left[\begin{array}{l} \text{MORSYN} \quad \boxed{1} \\ \text{MARKING} \quad none \\ \text{CONT} \quad \boxed{2} \\ \text{VAL} \quad \boxed{3} \end{array} \right] \right\rangle \end{array} \right]$$

References

- Ackerman, F and Stump, G. T. (2004). 'Paradigms and periphrastic expression'. In L. Sadler and A. Spencer (eds.), *Projecting Morphology*. Stanford: CSLI Publications, 111–157.
- Ackerman, F and Webelhuth, G. (1998). *A Theory of Predicates*. Stanford: CSLI Publications.
- Bresnan, J. (2001). *Lexical-functional syntax*. Oxford: Basil Blackwell.
- Gazdar, G., Klein, E., Pullum, G. K., and Sag, I. A. (1985). *Generalized Phrase Structure Grammar*. Harvard: Harvard University Press.
- Hinrichs, E. and Nakazawa, T. (1994). 'Linearizing AUXs in German verbal complexes'. In J. Nerbonne, K. Netter, and C. Pollard (eds.), *German in HPSG*. Stanford: CSLI Publications, 11–37.
- Jahani, C. (2000). 'Expressions of indirectivity in spoken Modern Persian'. In L. Johanson and B. Utas (eds.), *Evidentials: Turkic, Iranian and Neighbouring Languages*. Berlin: Mouton de Gruyter, 185–207.
- Lazard, G. (1985). 'L'inférentiel ou passé distancié en persan'. *Studia Iranica*, 14:27–42.
- Lazard, G., Richard, Y., Hechmati, R., and Samvelian, P. (2006). *Grammaire du persan contemporain*. Tehran: IFRI and Farhag-e Moaser.
- Poser, W. J. (1992). 'Blocking of phrasal constructions by lexical items'. In I. A. Sag and A. Szabolcsi (eds.), *Lexical matters*. Stanford: CSLI Publications, 111–130.
- Sag, I. A. (2007). 'Sign-based construction grammar: An informal synopsis'. Ms., Stanford University.
- Spencer, A. (2005). 'Inflecting clitics in generalized paradigm function morphology'. *Lingue e Linguaggio*, 4:179–193.
- Stump, G. T. (2001). *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.
- Windfuhr, G. (1982). 'The verbal category of inference in Persian'. In *Momentum Georg Morgenstierne II*. Leiden: E. G. Brill, 263–287.