Persian object clitics and the syntax-morphology interface

Pollet Samvelian
Université de Paris 3

Jesse Tseng
CNRS

Proceedings of the 17th International Conference on
Head-Driven Phrase Structure Grammar

Université Paris Diderot, Paris 7, France

Stefan Müller (Editor)

2010

Stanford, CA: CSLI Publications

pages 212–232

Samvelian, Pollet & Jesse Tseng. 2010. Persian object clitics and the syntax-morphology interface. In Stefan Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, Université Paris Diderot, Paris 7, France*, 212–232. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2010.12.

Abstract

This paper presents a descriptive overview and formal analysis of the use of pronominal clitics for realizing various types of arguments in Persian, with particular emphasis on object clitics in the verbal domain. We argue that pronominal clitics behave more like suffixes than independent syntactic elements; in cases where they take syntactic scope over an NP or a PP, they must be phrasal affixes. We propose an HPSG analysis to account for the morphosyntactic aspects of verbal suffixation of object clitics, possessive clitics, preverbal object clitics, and clitic doubling constructions. Finally, we explore extensions of the analysis to periphrastic verb forms, and we compare our proposals for Persian to previous HPSG work on clitic phenomena in other languages.

1 Introduction and data

1.1 Forms and functions

Persian has two sets of personal pronoun forms: full forms (1a) and enclitic forms (1b) (Lazard, 2006, §87, §91).[1]

(1) a. full forms:

	sg	pl
1	man	mâ(hâ)
2	to	šomâ(hâ)
3 (anim.)	u	išân (išun)
3 (inan.)	ân (un)	ânhâ (in(h)â)

b. enclitic forms:

	sg	pl
1	**-am**	**-emân** (**-emun**)
2	**-at** (**-et**)	**-etân** (**-etun**)
3	**-aš** (**-eš**)	**-ešân** (**-ešun**)

Full pronouns and enclitic pronouns can be used, often interchangeably, to express nominal arguments in a variety of constructions, but their morphosyntactic properties are highly divergent. We will consider two kinds of pronominal functions.

First, pronouns can be used to realize the nominal argument of a noun, adjective, or preposition:[2]

(2) adnominal argument (e.g. possessive):

[†]We wish to thank the participants of the HPSG seminar at Paris Diderot University, as well as the anonymous reviewers and participants of the 2010 HPSG conference. Special thanks go to Olivier Bonami, Philip Miller, François Mouret, and Gert Webelhuth. This work is supported by the bilateral project "PerGram", with funding from the ANR (France) and the DGfS (Germany) [grant no. MU 2822/3-I].

[1] Colloquial/familiar variants are shown in parentheses. With a few exceptions, the examples in this paper adopt literary/formal pronunciation.

[2]In addition to familiar categories (person/number, etc.), the following abbreviations are used in glosses: DDO = the definite direct object marker *râ*, EZ = the *ezafe* linking vowel *(y)e*, IPF = imperfective, SBJ = subjunctive.

pesar-e Maryam / pesar-e **u** / pesar-**aš**
son-EZ Maryam / son-EZ PRO.3SG / son-3SG

'Maryam's son / her son / her son'

(3) object of preposition:

barâ-ye Maryam / barâ-ye **u** / barâ-**yaš**
for-EZ Maryam / for-Z PRO.3SG / for-3SG

'for Maryam / for her / for her'

As we can see from these examples, full pronouns basically have the same syntactic distribution as NPs, like the proper noun *Maryam*.

Second, pronouns can be used to express an argument of a verb.[3]

(4) a. (mâ) Maryam-râ did-im / (mâ) **u**-râ did-im
we Maryam-DDO saw-1PL we PRO.3SG-DDO saw-1PL
'We saw Maryam.' / 'We saw him/her.'

b. (mâ) did-im-**aš**
we saw-1PL-3SG
'We saw her/him/it.'

Again, the full pronoun *u* has an NP-like distribution, very different from that of the enclitic *-aš*, which in this case is attached directly to the verb.

Clitic doubling is possible in colloquial registers. In other words, a single argument can be realized simultaneously as a syntactic complement (ordinary NP or full form pronoun) and as a clitic on the verb.

(5) Maryam-râ did-im-**aš** / u-râ did-im-**aš**
Maryam-DDO saw-1PL-3SG PRO.3SG-DDO saw-1PL-3SG
'We saw Maryam.' / 'We saw him/her.'

1.2 Preverbal object clitics

Instead of appearing with the verb as in the previous examples, object clitics can be realized on a variety of hosts to the left of the head verb. For example, Persian has a large number of compound predicates consisting of a lexical verb and a "preverb", typically a noun, adjective, or adverb that can be treated as a kind of grammaticalized complement. A direct object clitic can appear on either one of these elements:

(6) a. bâz kard-im-**aš**
open did-1PL-3SG
'We opened it.'

[3] See fn. 8 for the forms of the subject agreement markers (e.g. *-im*), which are not to be confused with the object clitics under discussion here.

b. bâz-**aš** kard-im
open-3SG did-1PL

An object clitic can also attach to a phrasal host, in most cases a PP:

(7) a. [$_{PP}$ ru-ye miz] gozâšt-im-**aš**
on-EZ table put-1PL-3SG
'We put it on the table.'

b. [$_{PP}$ ru-ye miz]-**aš** gozâšt-im
on-EZ table-3SG put-1PL

Clitics in preverbal position are sometimes ambiguous, allowing either an object clitic reading, or an adnominal clitic reading. For example, the PP in (7b) could instead be interpreted as a possessive: *ru-ye* [*miz-**aš***] 'on **his/her** table'.

Preverbal realization of object clitics is subject to various constraints. First, a single argument cannot be cliticized twice (as a preverbal clitic and as a clitic on the verb):

(8) *bâz-**aš** kard-im-**aš**
open-3SG did-1PL-3SG
(intended) 'We opened it.'

However, as we saw for clitics on the verb in (5), a preverbal clitic can double an NP object (in colloquial registers):

(9) a. <u>dar-râ</u> bâz-**aš** kard-im
door-DDO open-3SG did-1PL
'We opened the door.'

b. <u>ketâb-râ</u> [$_{PP}$ ru-ye miz]-**aš** gozâšt-im
book-DDO on-EZ table-3SG put-1PL
'We put the book on the table.'

Preverbal clitics are also sensitive to the syntactic function of their host. As we just saw in examples (6)–(7), they can attach to another complement of the verb. Adjuncts, on the other hand, cannot host object clitics:

(10) a. [$_{PP}$ dar xiâbân] did-im-**aš**
in street saw-1PL-3SG
'We saw him/her/it in the street.'

b. *$^{?}$ [$_{PP}$ dar xiâbân]-**aš** did-im
in street-3SG saw-1PL

(11) a. zud did-im-**aš**
early saw-1PL-3SG
'We saw him/her/it early'

b. * zud-**aš** did-im
early-3SG saw-1PL

Although they are attached to a host on their left, preverbal object clitics are also subject to a strong contextual constraint to their right: they must be immediately followed by the head verb. In the following example, the object clitic can attach to the preverb *nešân*, but not to the preceding PP complement:[4]

(12) a. (ketâb-hâ-râ) [be doxtar] nešân-**ešân** dâd-im
book-PL-DDO to girl show-3PL gave-1PL
'we showed them (the books) to the girl'

b. * (ketâb-hâ-râ) [be doxtar] **-ešân** nešân dâd-im
book-PL-DDO to girl -3PL show gave-1PL

Two clitic objects are possible in some ditransitive constructions, but they cannot appear on the same host. The only possibility in such cases is to have one preverbal clitic immediately before the verb, and one clitic on the verb (13d).

(13) a. ketâb-râ be to nešân dâd-im
book-DDO to PRO.2SG show gave-1PL
'We showed you the book.'

b. nešân *dâd-im-**at-aš** / *dâd-im-**aš-at**
show gave-1PL-2SG-3SG / gave-1PL-3SG-2SG

c. *nešân-**at-aš** / *nešân-**aš-at** dâd-im
show-2SG-3SG / show-3SG-2SG gave-1PL

d. nešân-**at** dâd-im-**aš** / nešân-**aš** dâd-im-**at**
show-2SG gave-1PL-3SG / show-3SG gave-1PL-2SG
'We showed it to you.'

As this previous example illustrates, beneficiary arguments can sometimes be realized as object clitics. This possibility is quite restricted, however, and it may be related to the fact that with some verbs, the beneficiary argument can be realized either as a *be*-PP as in (13a), or as an accusative NP (Lazard, 2006, §176.1). The constraints governing these alternations are not completely understood. We note furthermore that PP complements disallow clitic doubling:

(14) * ketâb [$_{\text{PP}}$ be to] nešan-**at** dâd-im / nešan dâd-im-**at**
book to PRO.2SG show-2SG gave-1PL / show gave-1PL-2SG
(intended) 'We showed a book to you.'

2 Arguments for affixal status

It is rarely straightforward to decide whether a clitic-host sequence should be analyzed syntactically or morphologically, because by definition, clitics present a

[4]Example (12b) is ungrammatical given the intended interpretation (indicated by the bracketing). The sentence is acceptable, however, with a possessive interpretation of the clitic: *be* [*doxtar-**ešan***] 'to **their** daughter'.

combination of word-like and affix-like properties. In this section, we will review a number of phonological and morphological facts that suggest strongly that pronominal enclitics in Persian are best analyzed as suffixes.

2.1 Phonological effects

Certain phonological adjustments can be observed when a vowel-initial pronominal clitic attaches to a vowel-final host. Some vowel sequences (e.g. *i-e*, *i-a*, *e-a*) are allowed (15a), but in other cases, the hiatus is broken by the insertion of the glide *y*:

(15) a. gorbe + **aš** → gorbe-**aš** 'his/her cat'
b. pâ + **-aš** → pâ-**yaš** 'his/her foot'
c. pâ + **-emân** → pâ-**yemân** 'our foot'

In colloquial Persian, the initial vowel of the clitic is often elided in such cases:[5]

(16) a. pâ + **-eš**, pâ + **-emun** → pâ-**š**, pâ-**mun** 'his/her foot, our foot'
b. did-i + **eš** → did-i-**š** 'saw-2SG-3SG' ⇝ 'you saw him/her/it'

Similar effects can be found with other clitics and at other morpheme boundaries. For example, glide insertion occurs before the *ezafe* linking vowel and before subject agreement markers.[6]

(17) a. xâne + **-e** → xâne-**ye** 'house-EZ'
b. mi-farmâ + **-ad** → mi-farmâ-**yad** 'IPF-order-3SG' ⇝ 'he orders'

In contrast, such effects are not observed at the boundary between two syntactic words. For example, there is no glide insertion between a preposition and its NP object:

(18) bâ âb / *bâ **y**-âb; tu âb / *tu **y**-âb 'with water; in the water'

While the foregoing examples show that pronominal clitics are more closely bound to their hosts than the elements in an ordinary syntactic combination, these facts are not wholly incompatible with a syntactic approach. A pronoun like *-aš* could be taken to be a syntactic word with a special marking like [+CLITIC] (to distinguish it from the full pronoun *u* 'he/she'). This marking could then license the phonological adjustments described above (vowel elision and glide insertion) as productive, "low-level" strategies for resolving hiatus.

This approach runs into difficulties, however, with the following data, involving prepositions. In colloquial Persian, some prepositions can combine with a clitic object, as we saw in (3) above.[7] The prepositions *be* and *bâ* exhibit unexpected

[5] For the pronunciation of the clitics, see fn. 1.

[6] See Lazard (2006, §22, §118).

[7] Those that cannot could be assumed, within a syntactic analysis, to subcategorize for a [−CLITIC] complement. This would account for contrasts like the following:

morphophonological effects with clitic objects. The initial vowel of the clitic can be elided (19a), just as in (16) above. Glide insertion, however, is not possible (19b); instead, we find idiosyncratic forms containing an inserted *h* (19c).

(19) a. be + **-eš**, bâ + **-emun** → be-**š**, bâ-**mun** 'to him, with us'
b. *be-**yeš** (*be-**aš**), *bâ-**yemun** (*bâ-**yemân**)
c. be-**heš**, bâ-**hâmun**

We could assume, following de Fouchécour (1981, p. 82), that these two prepositions have long forms *behe* and *bâhâ*, used exclusively with [+CLITIC] complements (while the forms *be* and *bâ* are compatible with all types of complements). But this would not explain why only vowel elision can apply to the resulting syntactic combinations, and not glide insertion. We prefer to analyze these preposition + clitic sequences as grammaticalized morphological compounds, for which such gaps and idiosyncrasies are more typical and can be dealt with in terms of familiar morphological notions such as allomorphy, suppletion, and defectivity.

2.2 Co-occurrence constraints

It is clear from the examples we have seen up to now that pronominal clitics allow "promiscuous attachment" to a wide range of hosts, in particular phrasal hosts. This could be taken as an argument in favor of syntactic combination. We will show in this section, however, that clitics are in fact sensitive to the lexical and morphological properties of their hosts, and that these facts cannot always be accounted for by syntactic means, such as subcategorization.

First of all, let us consider some cases that are potentially compatible with a syntactic approach. Participles, for example, can combine with a (possessive) pronominal clitic when used adjectivally (20a), but in verbal constructions they cannot host object clitics (20b):

(20) a. pirârhan-e šoste-**aš**
dress-EZ washed-3SG
'her washed dress'
b. * (pirâhan-râ) šoste-**aš**, va sepas ân-râ otu kard
dress-DDO washed-3SG, and then it-DDO iron did
'He/she washed the dress and then ironed it.'

Similarly, while we have seen many examples of object clitics attached to simple past tense and present tense verbs, present perfect forms do not allow this:[8]

(i) dar man / tâ man inside me, until me ([−CLITIC])
(ii) *dar-am / *tâ-yam inside me, until me ([+CLITIC])

[8]The present perfect involves a participial form followed by an enclitic form of the auxiliary *budan* 'be', which we assume, following Bonami and Samvelian (2009), to be a suffix. This auxiliary

(21) a. bâz kard-im-**aš** (= 6a)
open did-1PL-3SG
'We opened it.'

b. * bâz karde-im-**aš**
open done-1PL-3SG
(intended) 'We have opened it.'

The contrasts in (20)–(21) clearly cannot be explained phonologically. But the hosts involved do have distinct lexical representations, and so they could impose different constraints on the realization of their direct object: [±CLITIC] in the (a) examples, and [−CLITIC] in the (b) examples. Note, however, that the verb *karde-im* in (21b) does in fact allow a clitic object, if it is preverbal:

(21) c. bâz-**aš** karde-im
open-3SG done-1PL
'We have opened it.'

The syntactic analysis could still be saved, for example by introducing further features to distnguish clitics on the verb and preverbal clitics, but we prefer to treat the ungrammaticality of (20b) and (21b) as a morphological fact: pronominal clitics are suffixes, and the verb forms in these examples are simply incompatible with this type of suffixation.

Other systematic restrictions on pronominal enclisis present even more problems for the syntactic approach. As we saw above in (13c-d), there can be at most one pronominal clitic per host. This is true even if the clitics have distinct syntactic functions and scope. Compare, for example, sentence (7b), repeated here as (22a), and (22b), in which the PP complement happens to end with a possessive clitic:

(22) a. [$_{PP}$ ru-ye miz] **-aš** gozâšt-im (= 7b)
on-EZ table -3SG put-1PL
'We put it on the table.'

b. * [$_{PP}$ ru-ye miz-**at**] **-aš** gozâšt-im
on-EZ table-2SG -3SG put-1PL
(intended) 'We put it on **your** table.'

clitic is distinct from the subject agreement suffixes found with other verb forms, although the two paradigms are nearly identical:

(i) a. subject agreement suffixes:

	sg	pl
1	-am	-im
2	-i	-id (-in)
3	**-ad** (-e)	-and (-an)

b. enclitic auxiliary *budan*:

	sg	pl
1	-am	-im
2	-i	-id (-in)
3	**-ast** (-e)	-and (-an)

Note also that the 1sg form in both paradigms is identical to the 1sg object clitic, *-am* (1b). To avoid confusion, no examples with 1sg subjects are used in this paper.

Under a syntactic analysis, the clitic *-aš* combines with a PP in both cases, and given standard assumptions about locality, it should not be sensitive to the detailed morphological structure of a particular word within the PP. On the other hand, if *-aš* is a suffix, i.e. morphologically integrated into the right-most word of the host PP, then the contrast between *miz-**aš*** and **miz-**at-aš*** can be explained straightforwardly at the lexical level, by formulating restrictions on multiple suffixation.

Pronominal clitics also cannot co-occur with *ezafe*, which we have already encountered in several examples. This linking element, with the form *(y)e*, licenses the realization of NP-internal dependents to the right of the head noun. Following Samvelian (2007), we treat *ezafe* as a phrasal suffix. In example (23a), the noun *lebâs* must carry this suffix in order to combine with the adjective *sefid*, and the resulting phrase must be suffixed in order to combine with a possessive NP or full pronoun. In contrast, the second *ezafe* must not appear if the possessive pronoun is realized as a clitic (23b).

(23) a. lebâs-e sefid-**e** Maryam / lebâs-**e** sefid-**e** u
dress-EZ white-EZ Maryam / dress-EZ white-EZ PRO.3SG
'Maryam's white dress / her white dress'

b. lebâs-e *sefid-**e-yaš** / lebâs-e sefid-**aš**
dress-EZ white-EZ-3SG / dress-EZ white-3SG
'her white dress'

The fact that no *ezafe* appears on the adjective in (23b) indicates clearly that *-aš* is not a syntactic dependent within the NP. Instead, it is a suffix that attaches to the adjective morphologically (although, as a phrasal affix, it has syntactic and semantic scope over the whole NP).

Samvelian (2007) demonstrates that pronominal clitic + *ezafe* sequences are also excluded. In the following example, the relative clause must take *ezafe* to allow the realization of the genitive/possessive NP *in dâstân* 'of this novel' to the right. This is impossible in (24a), however, because the last word of the relative clause, *mihan-**aš*** 'his homeland', already carries a pronominal suffix:

(24) a. *qahremân-e [$_{\text{RC}}$ rânde šode az mihan-**aš**] **-e** in
hero-EZ driven become from homeland-3SG -EZ this
dâstân
novel
(intended) 'the hero of this novel, (who is) driven away from his homeland'

b. qahremân-e [$_{\text{RC}}$ az mihan-**aš** rânde šode] **-ye** in dâstân
hero-EZ from homeland-3SG driven become -EZ this novel

If the suffixed PP is moved away from the right edge of the relative clause, the incompatibility disappears, and the relative clause can receive the *ezafe* suffix (24b). Again, these facts would be difficult to analyze if *-aš* and *-(y)e* were syntactic elements, but they are readily explained if we assume that both forms are suffixes that cannot appear simultaneously on the same word.

2.3 Extraction

A last piece of evidence for the affixal status of pronominal clitics involves extraction. An object clitic must be fronted along with its host constituent (25b):

(25) a. mi-xâh-i fardâ bâz-**aš** bo-kon-i
IPF-want-2SG tomorrow open-3SG SBJ-do-2SG
'You want to open it tomorrow.'

b. bâz-**aš** [$_S$ agar mi-xâh-i fardâ ___ bo-kon-i]
open-3SG if IPF-want-2SG tomorrow SBJ-do-2SG
'If you want to open it tomorrow . . . '

c. * bâz [$_S$ agar mi-xâh-i fardâ ___-**(y)aš** bo-kon-i]
open if IPF-want-2SG tomorrow -3SG SBJ-do-2SG

d. * **-aš** [$_S$ agar mi-xâh-i fardâ bâz-___ bo-kon-i]
3SG if IPF-want-2SG tomorrow open SBJ-do-2SG

The clitic cannot simply be stranded and attach to a new host (25c).[9] And unsurprisingly, the clitic cannot be fronted without its host (25d). These facts are not wholly incompatible with an analysis of clitics as specially marked [+CLITIC] syntactic elements, with several additional assumptions and stipulations. But they follow automatically if *baz-**aš*** is analyzed as a single word (that is nevertheless interpreted as realizing two separate arguments of the verb).

In section 1.2, we stated that preverbal clitics had to immediately precede the verb; recall example (12). We can see now that this constraint is both too strong and too weak. Too strong, because the fronted clitic in (25b) is exempt from this constraint. Too weak, because the ungrammatical example (12b) remains ungrammatical even if the preverb *nešân* is extracted:

(26) * nešân [$_S$ agar ketâb-hâ-râ [be doxtar] **-ešân** ___ dâd-im]
show if book-PL-DDO to girl -3PL gave-1PL
(intended) 'if we showed the books to the girl'

The correct generalization appears to be, therefore, that preverbal clitics must be hosted by the least oblique complement of the verb, and that within the clause (i.e. if they are not extracted along with their host) they must appear immediately before the verb.

[9] Recall from (11) that preverbal clitics cannot attach to adverbial modifiers.

3 HPSG analysis of object clitics

3.1 Morphophonological functions

We adopt the insights of Miller and Sag (1997) in order to analyze the morphological realization of pronominal clitics as affixes. We briefly review the original analysis of French clitics, before presenting our proposed extension of the model to the Persian data.

Miller and Sag treat subject and object pronominal clitics in French as affixes on the verb. A sentence like *Je vous les donne* 'I give them to you' is thus analyzed as a single syntactic word, consisting of the finite verb *donne* and three pronominal affixes: *Je-vous-les-donne*.

The key technical device in their analysis is the morphophonological function $\mathbf{F}_{PRAF}$, which takes as input the inflected form of the verb (in I-FORM), its HEAD value (which determines prefixal vs. suffixal realization of pronouns), and its ARG-ST value. Elements on the ARG-ST list are typed as either *canonical-* or *affix-synsem* objects, and of course they carry grammatical specifications like the case and agreement features of each argument. Given this information, $\mathbf{F}_{PRAF}$ outputs the appropriate phonological form for the cliticized verb.

(27) *clitic-wd* →

$$\begin{bmatrix} \text{MORPH} & \begin{bmatrix} \text{FORM} & \mathbf{F}_{PRAF}(\boxed{0},\boxed{1},\boxed{2}) \\ \text{I-FORM} & \boxed{0} \end{bmatrix} \\ \text{SYNSEM} & \begin{bmatrix} \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{ARG-ST} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

For Persian, we propose a similar function, $\mathbf{F}_{pron}$, which requires four parameters instead of three. These include, of course, the I-FORM of the host and its ARG-ST list. The HEAD value is also necessary, not to determine the position of pronouns (unlike in French, Persian pronouns are always suffixed) but because $\mathbf{F}_{pron}$ is defined for both verbal and non-verbal hosts. Finally, the fourth parameter is the EDGE | RIGHT value, which contains the PRONARG feature, whose function will be explained in section 3.3 below.

(28)

$$\begin{bmatrix} \text{MORPH} & \begin{bmatrix} \text{FORM} & \mathbf{F}_{pron}(\boxed{1},\boxed{2},\boxed{3},\boxed{4}) \\ \text{I-FORM} & \boxed{1} \end{bmatrix} \\ \text{SSM} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{HEAD} & \boxed{2} \\ \text{ARG-ST} & \boxed{3} \\ \text{EDGE} \mid \text{R} & \boxed{4}\begin{bmatrix} \text{PRONARG} & \textit{index} \vee \textit{none} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

3.2 Suffix appearing on the verb

The following examples involve the inflected ditransitive verb *gozâšt-im* 'we put', for which we assume the following basic lexical description:[10]

(29) *gozâšt-im* 'put-1PL' $\rightsquigarrow$ 'we put'

$$\begin{bmatrix} \text{MORPH} & \begin{bmatrix} \text{FORM} & \mathbf{F}_{pron}(\boxed{1},\boxed{2},\boxed{3},\boxed{4}) \\ \text{I-FORM} & \boxed{1}\ \textit{gozâšt-im} \end{bmatrix} \\ \text{HEAD} & \boxed{2}\ \textit{verb} \\ \text{ARG-ST} & \boxed{3}\left\langle \text{NP}_{1pl},\ \text{NP}[\textit{acc}],\ \text{PP} \right\rangle \\ \text{EDGE} \mid \text{R} & \boxed{4}\left[\text{PRONARG} \quad \textit{none}\right] \end{bmatrix}$$

For our purposes, the I-FORM value can be a simple phonological string, but in actuality it contains a richer morphological representation.[11] In this description, the verb's accusative NP argument and its PP argument are underspecified, so the value of $\mathbf{F}_{pron}$ is as yet undetermined.

In the first example, the ARG-ST list in (29) is instantiated to require a canonical PP argument, but an NP argument of type *affix-synsem* with 3sg agreement features.

(30) *gozâšt-im-**aš*** 'put-1PL-3SG' $\rightsquigarrow$ 'we put it'

$$\begin{bmatrix} \text{FORM} & \mathbf{F}_{pron}(\textit{gozâšt-im}, \textit{verb}, \boxed{3}, [\text{PRONARG } \textit{none}]) = \textit{gozâšt-im-}\mathbf{aš} \\ \text{ARG-ST} & \boxed{3}\left\langle \text{NP}_{1pl},\ \text{NP}_{3sg}[\textit{aff}],\ \boxed{pp}\ \text{PP}[\textit{canon}] \right\rangle \\ \text{COMPS} & \left\langle \boxed{pp} \right\rangle \end{bmatrix}$$

Given an ARG-ST of this form as input, the effect of $\mathbf{F}_{pron}$ is to add the suffix *-aš* to the inflected verb. Following HPSG argument mapping principles, non-canonical *synsem* objects such as affixes are not mapped to the valence lists. In this case, the affix NP is not mapped to COMPS and therefore will not give rise to an additional, syntactic realization of the direct object. The PP argument, on the other hand, is mapped to COMPS and therefore realized canonically:

(31) [$_{\text{PP}}$ ru-ye miz] gozâšt-im-**aš**
on-EZ table put-1PL-3SG
'We put it on the table.'

Recall from example (5) above that clitic doubling is observed in colloquial Persian. To account for this, $\mathbf{F}_{pron}$ adds an optional pronominal suffix corresponding to a canonical argument:[12]

[10] As explained below in section 3.5, we further assume that all elements on ARG-ST in this basic (underived) lexical entry carry the feature [PRONARG *none*].

[11] See Bonami and Samvelian (2009) for a treatment of Persian verbal morphology using Paradigm Function Morphology within HPSG.

[12] As it stands, our formulation implies free variation between the presence and absence of the suffix. In reality, the stylistic effects associated with clitic doubling would need to be incorporated

(32) *gozâšt-im**(-aš)*** ‘put-1PL(-3SG)’ $\rightsquigarrow$ ‘we put’

$$\begin{bmatrix} \text{FORM} & \mathbf{F}_{pron}(gozâšt\text{-}im, verb, \boxed{3}, [\text{PARG } none]) = gozâšt\text{-}im\textbf{(-aš)} \\ \text{ARG-ST} & \boxed{3}\left\langle \text{NP}_{1pl}, \boxed{np}\ \text{NP}[canon], \boxed{pp}\ \text{PP}[canon] \right\rangle \\ \text{COMPS} & \left\langle \boxed{np}, \boxed{pp} \right\rangle \end{bmatrix}$$

In this case, the verb may be suffixed, but the NP argument is still mapped to COMPS and gives rise to the realization of a syntactic complement:

(33) ketâb-râ [$_{\text{PP}}$ ru-ye miz] gozâšt-im**(-aš)**
book-DDO on-EZ table put-1PL(-3SG)
‘We put the book on the table.’

3.3 Suffix appearing on a non-verbal host

Pronominal clitics can also attach to nouns and adjectives and some other non-verbal categories. In the general case, the host is a phrase, but in HPSG, syntactic phrases cannot undergo suffixation. A lexicalist analysis of phrasal affixation is possible, though, if we separate the morphological effects of the suffix (at the lexical level) and its syntactic and semantic effects (at the phrasal level).

The morphological realization of clitics on non-verbal hosts is exactly the same as in the case of verbal suffixation, so it is handled by the same function $\mathbf{F}_{pron}$. The following example illustrates the suffixation of the 3sg suffix *-aš* to the adjective *sefid* ‘white’:

(34) *sefid-**aš*** ‘white-3SG’

$$\begin{bmatrix} \text{MORPH} & \begin{bmatrix} \text{FORM} & \mathbf{F}_{pron}(\boxed{1}, \boxed{2}, \boxed{3}, \boxed{4}) \\ \text{I-FORM} & \boxed{1}\ sefid \end{bmatrix} \\ \text{HEAD} & \boxed{2}\ adj \\ \text{ARG-ST} & \boxed{3}\ \langle\ \rangle \\ \text{EDGE} \mid \text{R} & \boxed{4} \end{bmatrix}$$

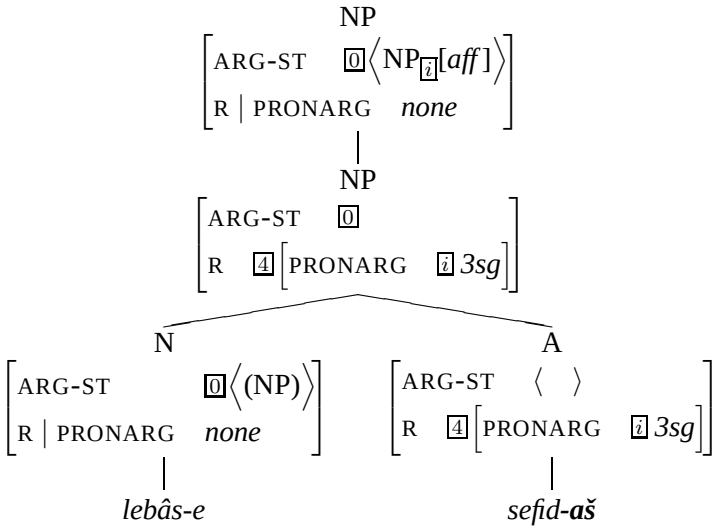
$$\sqsubset \begin{bmatrix} \text{MORPH} & \left[\text{FORM} \quad \mathbf{F}_{pron}(\boxed{1}, \boxed{2}, \boxed{3}, \boxed{4}) = sefid\textbf{-aš}\right] \\ \text{EDGE} \mid \text{R} & \boxed{4}\left[\text{PRONARG} \quad \begin{bmatrix} \text{PER} & 3rd \\ \text{NUM} & sg \end{bmatrix}\right] \end{bmatrix}$$

Unlike the examples in the previous section, $\mathbf{F}_{pron}$ does not constrain the host’s ARG-ST list (which in this case is empty). The only constraint that $\mathbf{F}_{pron}$ imposes is that the presence of the suffix (i.e. its 3sg index) must be recorded in PRONARG. We introduce this feature to handle the mismatch between the morphological scope of the suffix (a single word) and its syntactic/semantic scope (a phrase or clause).

into the grammatical description and added as an additional parameter to $\mathbf{F}_{pron}$.

To see how this works, consider our analysis of example (23b) above.[13]

(35) $[_{\text{NP}}$ *lebâs-e sefid*$]$ ***-aš*** 'dress-EZ white-3SG' $\rightsquigarrow$ 'her white dress'



As we have just seen, the suffixed adjective ***sefid-aš*** has a non-empty PRONARG value, but at the lexical level, the interpretation of this 3sg index is not yet determined. The common noun *lebâs* has an optional NP argument on its ARG-ST list (linked to a possessive relation in its semantic content), which is also uninstantiated at the lexical level. These two pieces of information can only be associated when the entire phrase ***lebâs-e sefid-aš*** is constructed.

This is why we defined PRONARG as a right edge feature. In branching phrases, the value of EDGE | RIGHT is shared between the rightmost daughter and the mother. We further assume that ARG-ST propagates as a HEAD feature. The result of this sharing of information can be seen in (35), where the relevant specifications are accessible when the head-adjunct phrase is formed. At this point, we can apply a unary syntactic rule that establishes the link between the PRONARG index and the possessive NP argument, and that also "discharges" the PRONARG value.

3.4 Preverbal object clitics

The PRONARG feature is also crucial in our analysis of the preverbal object clitics presented in section 1.2. In these cases, the clitic is again suffixed to the rightmost word of a phrase, but instead of realizing an argument of that phrase (like the possessive in the previous example), a preverbal object clitic must be interpreted at the level of the whole clause.

Example (7b), repeated here as (36a), contains a preverbal object clitic attached to a PP. Example (36b) involves the same structure, but with clitic doubling.

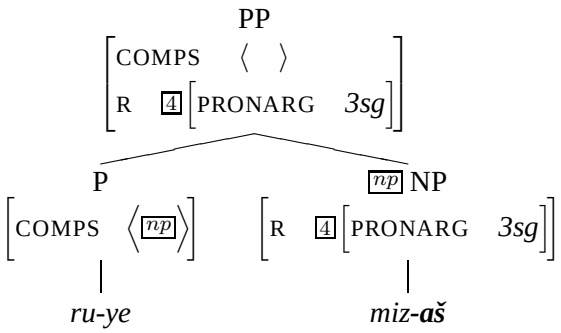
[13] We leave aside the analysis of the *ezafe* suffix in this example. We return briefly to the issue of *ezafe* in section 3.5, but for a full discussion, see Samvelian (2007).

(36) a. [$_{\text{PP}}$ ru-ye miz] **-aš** gozâšt-im
on-EZ table -3SG put-1PL
'We put it on the table.'

b. ketâb-râ [$_{\text{PP}}$ ru-ye miz] **-aš** gozâšt-im
book-DDO on-EZ table -3SG put-1PL
'We put the book on the table.'

The following figure shows the analysis of the suffixed PP complement found in these sentences:[14]

(37) [$_{\text{PP}}$ *ru-ye miz*] ***-aš*** $\rightsquigarrow$ 'on the table' + uninterpreted 3sg pronoun



Just as in (34) above, $\mathbf{F}_{pron}$ adds a suffix to the noun *miz* and the corresponding index becomes the value of the PRONARG attribute. This PRONARG value could be discharged at the NP level as in the previous section, giving rise to a possessive interpretation ('on his/her table'), but instead, in this case PRONARG continues to propagate to the level of the PP, where it remains uninterpreted.

To complete the analysis of the sentences in (36), we need to modify the verb *gozâšt-im* 'we put' so that it can accept the suffixed PP in (37) as its complement, as opposed to the ordinary PP that we saw in earlier examples like (31) and (33). We propose the following lexical rule:

(38) $$\begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{ARG-ST} & \boxed{1}\left\langle \ldots, \text{NP}_{\boxed{i}}[\textit{acc}], \ldots \right\rangle \oplus \left\langle \boxed{2}\begin{bmatrix}\text{PRONARG} & \textit{none}\end{bmatrix}\right\rangle \end{bmatrix}$$
$$\mapsto \begin{bmatrix} \text{ARG-ST} & \boxed{1} \oplus \left\langle \boxed{2'}\begin{bmatrix}\textit{non-aff} \\ \text{PRONARG} & \boxed{i}\end{bmatrix}\right\rangle \end{bmatrix}$$
where 2 and 2′ are identical except for their PRONARG values

The effect of this rule is to add the index of an accusative NP argument to the PRONARG value of the last element of ARG-ST, which corresponds to the least oblique argument. This argument thus becomes the clitic host, and it must not

[14] *Ru-ye* is in fact a grammaticalized nominal element with the *ezafe* suffix, but here we analyze it simply as a preposition.

itself be cliticized. The specification *non-aff* is compatible with either canonical realization or extraction (*gap-synsem*).

The change from [PRONARG *none*] to [PRONARG *index*] on the host argument ensures that the rule can only apply once: There can be only one preverbal clitic per clause. On the other hand, the original accusative NP remains on ARG-ST and its description is not further specified or modified in any way.

In the case of *gozâst-im*, the output of applying rule (38) to the basic lexical entry in (29) is as follows:

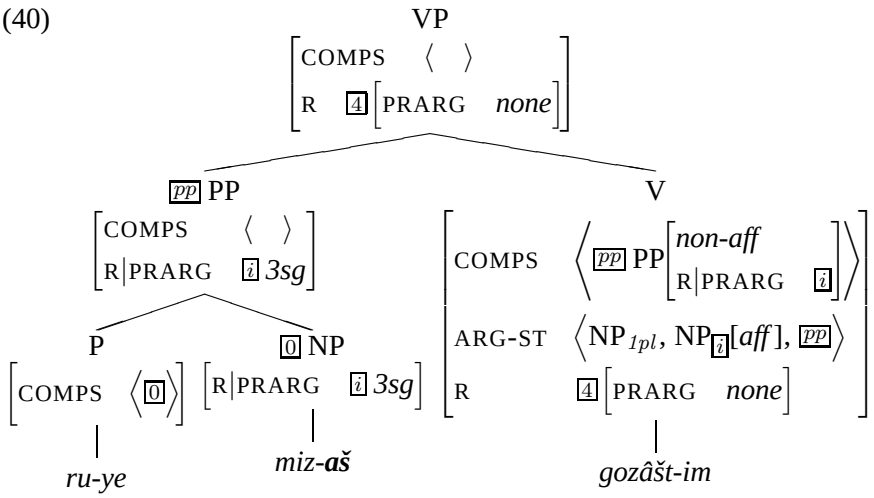
(39) *gozâšt-im* 'put-1PL' $\rightsquigarrow$ 'we put'

$$\begin{bmatrix} \text{FORM} & \mathbf{F}_{pron} = \textit{gozâst-im} \\ \text{HEAD} & \textit{verb} \\ \text{ARG-ST} & \left\langle \text{NP}_{1pl},\ \text{NP}_{\boxed{i}}[\textit{acc}],\ \text{PP}\begin{bmatrix}\textit{non-aff} \\ \text{PRONARG} \quad \boxed{i}\end{bmatrix} \right\rangle \end{bmatrix}$$

As indicated, the morphophonological function $\mathbf{F}_{pron}$ does not add a pronominal suffix to the verb if the corresponding index appears in the PRONARG value of an ARG-ST element.

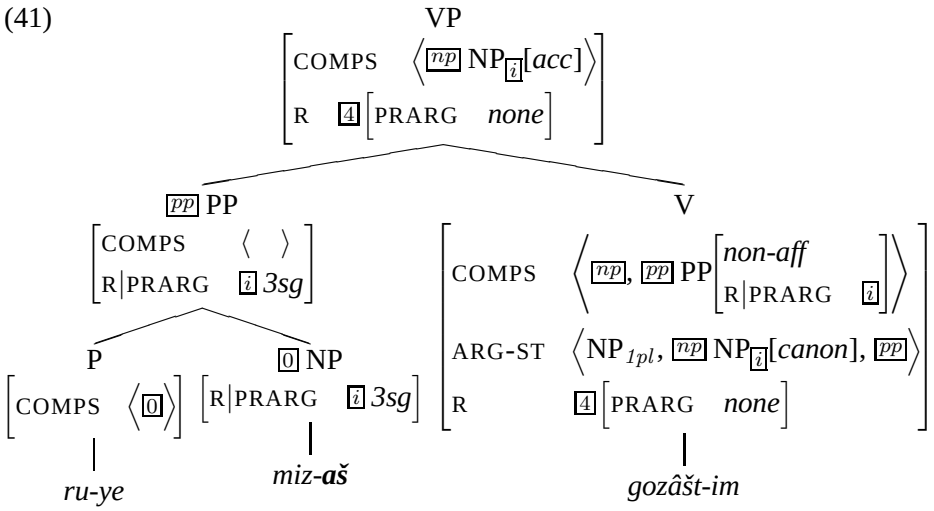
The accusative NP can be further instantiated as either affixal or canonical. In the first case, it is not mapped to COMPS, and the argument is only realized once, as in (36a), which we analyze as follows:

(40)



On the other hand, the accusative NP in (39) can be instantiated as canonical, giving rise to clitic doubling, as in example (36b), with the following (partial) analysis:

(41)



Unlike in (40), in this derivation the VP is not saturated, so it can still combine with the syntactic NP[*acc*] complement corresponding to the clitic *-aš*.

3.5 Remaining details

In this section we fill in a few remaining gaps in our formal analysis.

First, we assume that verbs (and other heads) are lexically specified as having only [PRONARG *none*] arguments:

(42) $\textit{lexeme} \rightarrow \left[\text{ARG-ST} \quad \textit{list}\left(\left[\text{PRONARG} \quad \textit{none}\right]\right)\right]$

Without this constraint, spurious object clitic pronouns, not corresponding to any argument, could be freely instantiated:

(43) * dar-râ bâz-**at** kard-im
door-DDO open-2SG did-1PL
'We opened the door.' + uninterpreted 2sg pronoun

With (42) in place, unless the verb *kard-im* explicitly undergoes a derivational process like the lexical rule in (38), its complement *bâz* cannot host a preverbal clitic.

The fact that (38) only applies to arguments of the verb accounts for the ungrammaticality of adjunct hosts, as illustrated in (10)–(11).

The various clitic co-occurrence constraints discussed in section (2.2) are handled by $\mathbf{F}_{pron}$. For example, multiply-suffixed forms like **dâd-im-**aš-at*** in (13b) and **miz-at-aš*** in (22b) are simply never produced by $\mathbf{F}_{pron}$, no matter what the input. The incompatibility of clitic pronouns and *ezafe* can be accounted for because $\mathbf{F}_{pron}$ has access to all of the right edge features of the host. Since *ezafe* is a phrasal affix, there must be a corresponding (boolean) feature EDGE | RIGHT | EZ that encodes its presence. $\mathbf{F}_{pron}$ will only add a pronominal suffix to a host that carries the specification [−EZ] (absence of *ezafe*), and similarly, the morphophonological function $\mathbf{F}_{ez}$ that realizes *ezafe* requires its host to have the feature [PRONARG *none*].

As the final ingredient of our formal analysis, we need to formulate a linear precedence constraint to ensure that preverbal clitics appear immediately before the verb. Although the lexical rule (38) ensures that the host is the least oblique argument, we must still prevent modifiers and other intervening elements from appearing in the syntactic realization of the clause. The following LP rule requires the clitic host (i.e. any complement with a non-empty PRONARG specification) to immediately precede the head verb:

(44) $\begin{bmatrix}\text{COMP-DTR}\\ \text{PRONARG}\quad \textit{index}\end{bmatrix} \ll \begin{matrix}\text{HD-DTR}\\ \text{V}\end{matrix}$

This constraint specifies the grammatical functions of the elements involved. As a result, it correctly applies in head-complement phrases such as (12), but does not exclude head-filler phrases like (25).

Finally, we saw at the end of section 1.2 that some beneficiary arguments can also be realized as clitics. The definition of $\mathbf{F}_{pron}$ and the formulation of the lexical rule in (38) can be modified to accommodate the examples in (13), with an additional constraint on clitic doubling to account for (14). However, a more thorough empirical investigation is required before beneficiary arguments can be fully incorporated into our formal analysis.

4 Further questions and discussion

4.1 Clitics in periphrastic constructions

Thus far, our analysis of object clitics only covers clauses containing a single, simple verb form. Persian also has a variety of periphrastic verb forms, with highly divergent properties. A descriptive overview and an HPSG analysis of these constructions can be found in Bonami and Samvelian (2009). It remains to be seen whether the present proposals can be extended in harmony with that account.

The periphrastic constructions include the passive voice and several compound tenses, and they vary with respect to the following properties: the relative order of the finite auxiliary and the lexical verb, the morphosyntactic status of the auxiliary element (word or affix), the morphological form of the lexical verb (finite or non-finite/participial), and finally (and most importantly for us) the realization and placement of object clitics.

We already saw an example of a compound tense, the present perfect ("compound present" in the terminology of Bonami and Samvelian), in example (21) in section 2.2. In this tense, the auxiliary verb *budan* is realized as a suffix on the participle; in other words, the present perfect is not truly periphrastic. The resulting suffixed form is incompatible with further object clitic suffixation. This type of incompatibility can be integrated into the definition of $\mathbf{F}_{pron}$, which has access to the HEAD features of the verb (in particular, VFORM). Note that this restriction has no effect on the preverbal clitic in (21c), which is still correctly licensed by lexical rule (38).

The following examples illustrate the past perfect ("complex bounded past") tense, which involves a full form of the auxiliary *budan*, to the right of the participle. The auxiliary can host an object clitic (45a), but the participle cannot (45b).

(45) a. bâz karde bud-im-**aš**
open done was-1PL-3SG
'We had opened it.'

b. * bâz karde-**aš** bud-im
open done-3SG was-1PL

c. bâz-**aš** karde bud-im
open-3SG done was-1PL

The clitic on the head verb *bud-im* in (45a) and the preverbal clitic in (45c) are handled correctly by our analysis. To block the realization of the preverbal clitic in (45b), we assume that the participle *karde* is disqualified as a clitic host in the definition of $\mathbf{F}_{pron}$ (again via the HEAD | VFORM specification). We saw the effects of this morphological restriction on this same participial form in a different syntactic context in example (20b) in section 2.2.

Finally, we consider the future tense, which is the only compound tense where a non-finite lexical form appears to the right of the finite auxiliary. It is also the only construction where both the auxiliary and the lexical verb can host an object clitic:

(46) a. be Maryam xâh-im dâd-**aš**
to Maryam want-1PL give-3SG

b. be Maryam xâh-im-**aš** dâd
to Maryam want-1PL-3SG give
'We'll give it to Maryam.'

Bonami and Samvelian (2009) treat *xâh-im dâd* as a single inflected form. At first glance the clitic placement in (46b) seems problematic for this analysis, but in fact, since $\mathbf{F}_{pron}$ has access to the internal morphological structure of this verb form (encoded in the I-FORM value), it can be defined to realize the clitic *-aš* either as a suffix or as an infix.

While this approach is technically feasible, there appears to be no additional motivation for allowing infixation in the morphology of Persian. For this and other reasons (e.g. word order facts not taken into account by Bonami and Samvelian), it is useful to explore alternative, syntactic analyses of the future tense. We note some parallels between this structure and impersonal modal constructions:

(47) a. (u-râ) mi-tavân did-**aš**
PRO.3SG-DDO IPF-can saw-3SG

b. (u-râ) mi-tavân-**aš** did
PRO.3SG-DDO IPF-can-3SG saw
'One can see him/her.'

The "downstairs" lexical verb appears in the same bare stem form as in the future tense, and it can take an object clitic in the usual way, through head suffixation (47a). The clitic in (47b) cannot be analyzed as a preverbal clitic using the lexical rule in (38), because the modal is not a complement of *did*. On the contrary, *did* is a complement of the "upstairs" modal, and so (47b) is an instance of clitic climbing, for which we adapt the argument composition analysis proposed for related phenomena in Romance (Abeillé and Godard, 2002). We suggest following a similar approach for the future tense data in (46).

4.2 Cross-linguistic considerations

Similar cliticization phenomena are found in other Western Iranian languages. Sorani Kurdish, for example, also has preverbal object clitics. In fact, as the following examples from Bonami and Samvelian (2008) show, preverbal placement is the only possibility:

(48) a. min $[_{\mathrm{PP}}$ ba Narmîn] **-î** da-lê-m
I to Narmîn 3SG IPF-tell-1SG
'I am telling it to Narmin.'

b. * min $[_{\mathrm{PP}}$ ba Narmîn] da-lê-m-**î**
I to Narmîn IPF-tell-1SG-3SG

Our analysis of Persian can be easily adapted to account for this data.

Pronominal clitics are of course also found in many other language families. We already mentioned French pronominal clitics in section 3. More generally, pronouns in the Romance languages exhibit many of the same phenomena observed in Persian: the existence of weak (clitic) forms and strong forms, the affixal status of clitic forms used to realize the arguments of a verb, limited mobility (e.g. clitic climbing), and clitic doubling.

There are differences: unlike in Persian, Romance object clitics generally are not also used to realize dependents within the NP, Romance exhibits proclisis in addition to enclisis, and subject pronouns can also have clitic realization in Romance. In spite of these differences, there seems to be a rich common ground for comparative studies from a formal perspective.

As discussed in section 3.1, our analysis of Persian is inspired by Miller and Sag (1997), and we hope that further work (in particular on clitics in multi-verb structures) will be able to draw on existing HPSG analyses of Romance, and also provide new insights and develop analytical tools to improve upon earlier work.

Clitic phenomena in the Slavic languages have also received attention in HPSG in recent years, and should also be taken into account within this formal comparative perspective. A particularly striking parallel can be observed in the "floating" auxiliary clitics in Polish analyzed by Kupść and Tseng (2005). Much like Persian object clitics, these auxiliary clitics can appear either suffixed to the verb (49a), or attached to a dependent phrase to the left of the verb:

(49) a. Dlaczego [tak długo] nie pisała **-ś**?
why so long NEG written.FSG -2SG
'Why haven't you written in such a long time?'

b. Dlaczego [tak długo]-**ś** nie pisała?

c. Dlaczego-**ś** [tak długo] nie pisała?

The HPSG analyses proposed for Polish and Persian have very little in common in fact, primarily because auxiliaries and objects have completely different argumental properties. Nevertheless, the remaining morphosyntactic aspects of the analyses of the two languages, specifically concerning the constraints on the position of clitics within the clause, could be brought closer together.

We believe that existing analyses of clitic phenomena, such as those mentioned here, are now available in sufficient number to allow the development of a more general theory of clitics in HPSG. These efforts will provide a formal framework for typological research and guide us in the study of the many clitic phenomena, in Persian and in other languages, that await description and formal analysis.

References

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

Bonami, O. and Samvelian, P. (2008). Sorani kurdish person markers and the typology of agreement. 13th International Morphology Meeting, Vienna.

Bonami, O. and Samvelian, P. (2009). Inflectional periphrasis in Persian. In S. Müller, editor, *The Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar*, pages 26–46, Stanford. CSLI Publications.

de Fouchécour, C. (1981). *Éléments de persan*. Publications Orientalistes de France, Paris.

Kupść, A. and Tseng, J. (2005). A new HPSG approach to Polish auxiliary constructions. In S. Müller, editor, *Proceedings of the 12th International Conference on HPSG*, pages 253–273. CSLI Publications, Stanford, CA.

Lazard, G. (2006). *Grammaire du persan contemporain*. Institut Français de Recherche en Iran, Téhéran. Nouvelle édition avec la collaboration de Y. Richard, R. Hechmati et P. Samvelian.

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

Samvelian, P. (2007). A phrasal affix analysis of the Persian *Ezafe*. *Journal of Linguistics*, **43**(3), 605–645.