INCORPORATION AND COMPLEX PREDICATION IN PERSIAN

Fatemeh Nemati

University of Tehran, University of Konstanz

Proceedings of the LFG10 Conference

Miriam Butt and Tracy Holloway King (Editors)

2010

CSLI Publications

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

Abstract

This study deals with the nature of N+V sequences in Persian and suggests a sub-classification of these sequences into Noun Incorporation and Complex Predication. This classification is grounded in the lexical and phrasal properties of the nouns involved in these sequences. Noun Incorporation cases are analyzed in terms of head-adjunction of the non-projecting Noun at the level of c-structure (Toivonen 2001). Complex Predication is dealt with in terms of the predicate composition proposed in Butt (1995, 1997) and Alsina (1997), along with some adaptations from Pustejovsky's (1995) theory of the generative lexicon.

1 Introduction

Persian shows a strong preference for using multiword verbal expressions over simple verbs[1]. Sadeghi (1993) has stated the number of simple verbs used in both spoken and written Persian do not exceed 150. He also claims that this is not a new tendency and that the formation of multiword verbal expressions has been used extensively even before the enormous borrowings from Arabic, and later from other foreign languages. Therefore, Persian must have been using productive processes to conceptualize new ideas and add new verbs to its repository of verbal expressions. Since these processes have been used over centuries, it is no surprise if a once-productive process of verb formation is not accessible to the Persian speakers anymore (locative incorporation as in *piS raftan* 'front going' (to move forward)); and if the idiomaticity of some of the verbal expressions obscures their internal structures (consider *zamin xordan* 'ground hitting' (to fall) where object of proposition is incorporated by the verb). This short article would surely do not do justice to reflect upon the whole process and, thus, the focus of the present study will be limited to a synchronic investigation of frequent multiword-verbal expressions, mostly labeled as complex verbs, composite verbs or compound verbs in the state-of-the-art.

A close look at the constructions shows that these verbal expressions vary systematically from each other in terms of the semantic relation of the non-verbal element to the verbal element. They also show different syntactic behaviors. This study aims at investigating the nature of these multiword

1 I am grateful to Miriam Butt for her support and long discussions during my research period at the University of Konstanz. My thanks are also due to Ali Ashraf Sadeghi for his insightful seminars on Persian morphology. I would like to show my gratitude towards Yehuda Falk, George Aoran Broadwell, Lachlan Duncan, Louisa Sadler and Rachel Nordlinger for their helpful comments. I am also very grateful to my colleagues at the University of Konstanz, Thomas Mayer, Michael Spagnol, Maialen Iraola Azpiroz and my colleagues in Iran, Fatemeh Alavi and Sahar Bahrami for helping me at different stages of the work.

verbal expressions based on these two factors and will show that there are two different processes involved in the formation of verbal expressions, namely incorporation vs. complex predication. In the next section, I will touch on the differences of these two subclasses of multiword verbal expressions, limiting the scope of it to N+V sequences. In section 3 and 4, an attempt will be made to analyze these two types of multiword verbal expressions in the light of LFG findings and to show how they can be represented linguistically.

2 Differences between Multiword Verbal Expressions: Motivations for a sub-classification

2.1 The Basic Data

To start with the differences between multiword verbs in Persian, a very brief introduction to the sentence structure in Persian seems to be due. Persian is an SOV, pro-drop language with a partially free word order, to the extent that Sadeghi (pc) claims that it is a non-configurational language. The canonical word order is illustrated in (1a). Other possible word orders are given in (1b, c, d, and e).

(1) a. [āryā] [ketāb rā] [be man] [dād]
Arya book OM to me give.Past.3.Sg.
Subj OBJ OBL V
'Arya gave me the book.'
b. [ketāb rā] [āryā] [be man] [dād]
c. [be man] [āryā] [ketāb rā] [dād]
d. [āryā] [be man] [ketāb rā] [dād]
e. [ketāb rā] [be man] [āryā] [dād]

As can be seen from the above sentences, the canonical position of the verb is sentence final. The verb rarely undergoes scrambling. The object appears obligatory followed by 'rā' as the accusative case marker, if definite, as in the above sentences. Otherwise, it can be followed by an indefinite clitic (2), having the same ordering possibilities as a definite object.

(2) [āryā] [ketāb-i] [be man] [dād]
Arya book-INDF to me give.Past.3.Sg.
Subj OBJ OBL V
'Arya gave me a book.'

The OBL(ique) or indirect object always appears as the complement of a preposition and receives case as the object of preposition. The only noun in the sentence with a covert case is the subject, receiving a nominative case.

Keeping these characteristics of Persian sentence grammar in mind, we proceed to our topic: multiword verbal expressions. As an example of such constructions, consider the example sentences (3b) and (3c) compared to (3a).

(3) a. minā **Gazā** rā be bače **dād**
Mina food OM to child give.Past.3.sg
'Mina gave the food to the child.'

b. minā be bače **Gazā** **dād** (Noun Incorporation)
Mina to child food give.Past.3.sg
Lit. 'Mina food-gave to the child.'

c. minā āryā rā **šekast** **dād** (Complex Predication)
Mina Arya OM defeat give.Past.3.sg
'Mina defeated Arya.'

The verbal in all the three sentences is *dād*. (3a) and (3b) have an NP with the same semantic content (*Gazā*). They differ in that the noun in (3a) is followed by a case marker, giving it a specific interpretation, whereas the noun in (3b) appears caseless and adjacent to the verb. There are strict constraints on the interpretation and the order of this caseless noun and it also varies from the noun in the (3a) in having a generic interpretation and joining the verb to denote a unitary activity. I argue that this N+V sequence in (3b) is a case of Noun Incorporation (NI).

The N+V sequence in (3c) will, in contrast to (3b), be discussed as a Complex Predicate (CPr). The noun here appears caseless as well, but it has no argument relation to the verb, as can be inferred from the semantic content of the noun. The noun is part of the predication and contributes to the argument structure of the complex. What is predicated in (3c) does not refer to the main semantic content of the verb 'the transference of something from Arya to Mina', but that Arya has brought something on Mina that is 'defeat'. Besides this semantic difference, I point to other differences between these two N+V sequences in the remainder of the paper, arguing that they should receive different syntactic analyses.

2.2 Syntactic Behavior of CPrs vs. NI

2.2.1 Modification

The noun in incorporation sentences resists modification by adjectives and quantifiers (4a). Otherwise, modification changes the semantic interpretation of the complex as a conceptual whole to its non-incorporated counterpart, as it is can be seen in (4b). On the other hand, when the noun in CPr is modified (as in (4c)), the scope of the modification is extended over the whole event denoted by the N+V sequence and the semantics of the noun, in terms of

definiteness, does not change. This result is predicted because the noun in (4c) is an essential part of the predication.

(4) a. *minā koll-e **nāme** nevešt
Mina all-of letter write.Past.3.Sg
Lit. '*Mina the whole letter-wrote.'

b. minā **nāme**-'i *tulāni* nevešt
Mina letter-INDF long write. Past.3.Sg
'Mina wrote a long letter.'

c. (ānhā) dar gilān šekast-e *sangin*-i xord-and
(they) in Gilan defeat-Ez hard-INDF eat- Past.3.Pl
'They were defeated severly in Gilan'.

2.2.2 Relativization

The noun in the syntactic paraphrase of *Gazā xord-im* (4a) can be relativized and thus it becomes specific, referring to a certain instance of food (consider 4b). Relativization of the nominal in CPrs does not result in specific reading of the noun: the noun fills the subject position of the main clause, but the whole event predicated jointly by the noun and the verb is inferred to be functioning as the subject.

(5) a. diruz tu resturān Gazā xord-im.
Yesterday in restaurant food eat-Past.1.Pl.
Lit. 'We food-ate in the restaurant yesterday.'

b. Gazā-'i ke diruz xord-im xošmaze bud
food-INDF that yesterday eat-Past.1.Pl delicious be.Past.3.Sg.
'The food that we ate yesterday, was delicious.'

c. šekast-i ke mā xord-im be dalil-e
defeat-INDF REL we eat- Past.1.Pl to reason-EZ
na-dāštan-e barnāme bud
not-having-Ez plan be- Past.3.Sg
'We were defeated because of not having a plan.'

2.2.3 Scrambling

As Persian is a partially free word order language, the constituents might appear in almost any order before the verb (in the spoken register some constituents might appear after the verb as well). A bare noun with a generic interpretation as observed in (6a) cannot scramble. When the noun scrambles, it is modified and obtains a specific reading (6b). Scrambling of the nominal element of the CPr does not result in the specific interpretation of the noun.

Rather, it gives the event a pragmatic prominence, putting it in a focus position (6c).

(6) a. minā tamām-e ruz rā xub dars xānd
Mina all-Ez day OM well lesson read.Past.3.sg.
'Mina studied well all day long.'

b. minā dars-hā-yaš rā xub xānd
Mina lesson-Pl.-her OM well read.Past.3.sg.
'Mina studied her lessons well.'

c. diruz kotak-e šadidi-i āryā az bābā-š xord
yesterday beating-EZ harsh-a Arya from father-PossC eat.past.3.sg.
'Arya was beaten harshly by his father yesterday.'

2.2.4 Pronominal Cliticization

In Persian, clitics attach to the outer edge of phrasal constituents (7a) and (7b); they do not have access to the internal structure of words (7c).

(7) a. ketāb-aš
book-PosCl3.sg.
' her/his book'

b. ketāb-xāne-aš
book-house-PosCl3.sg.
'her/his/its library'

c. *ketāb-aš-xāne
book-PosCl3.sg.-house
'her/his/its library'

NI disallows cliticization which follows from the lexical status of the Noun. When a clitic attaches a noun in its corresponding non-incorporated paraphrase, the noun becomes specific (compare (8a) and (8b)). The nominal in CPr, however, allows cliticization, the clitic has no effect on the nonspecific interpretation of the noun (consider (8c) and (8d)). It is also worth mentioning that the type of clitics attaching the noun in the syntactic counterpart of incorporation is different from the clitic attaching the nominal in CPrs: in the former, it is a possessive clitic; while in the latter, it is a pronominal clitic satisfying one of the grammatical functions in the sentence (compare (8b) with (8d) and (8f) for the difference).

(8) a. az bānk **vām** **gereft** (NI)
from bank loan take.Past.3.sg
'S/He got a loan from the bank.'

b. **vam**-aš rā az bānk **gereft**
loan-POSCl.3.sg. OM from bank take.Past.3.sg
Lit. 'S/He got her/his loan from the bank.'

c. minā nasrin rā be mehmāni **da'vat** **kard** (CPr)
Mina Nasrin OM to party invitation do-Past.3.sg
'Mina inviter her/him to the party.'

d. minā be mehmuni **da'vat**-eš **kard** (spoken register, CPr)
Mina to party invitation-PCl3.sg. do-Past.3.sg
'Mina inviter her/him to the party.'

e. minā be nasrin **komak** **kard** (CPr)
Mina to Nasrin help do- Past.3.sg
'Mina helped Nasrin.'

f. minā **komak**-eš **kard** (spoken register, CPr)
Mina help-PCl3.sg. do-Past.3.sg
'Mina helped her.'

The fact that the noun in NI is invisible to syntactic processes shows that it has a lexical status. CPrs, in contrast, are syntactic and the noun has a phrasal status in that it can function as a host to clitics and it can be modified, relativized and scrambled.

2.3 An Overview of N+V Sequences in the State-of-the art

The investigation of Persian N+V sequences in the state-of-the-art has opted for either a lexical or a syntactic approach. Some researchers claim that all multiword verbal expressions are lexical and that they are the result of the morphological processes of incorporation and combination (Dabir-Moghaddam 1997, Vahedi-Langrudi 1996). However, a lexicalist approach falls short of explaining the syntactic behavior of CPrs, as discussed above.

Other researchers discuss multiword verbal expressions as syntactic constructions, but they fail to observe the distinction between the two types of the nouns in N+V sequences (Karimi 1997, Karimi-Doostan 1997, Folli et al. 2005, Pantcheva 2010). Mostly they have ignored the possibility of an incorporation account for some of the N+V sequences, as well as ADV+V

sequences whose resistance to separation can be explained better in the light of an incorporation account. Megerdoomian (2006), in contrast, deals with the syntactic and semantic differences between these two N+V sequences, treating the nominal part under the term bare nominal as opposed to preverbal nominal (in my analysis the incorporated noun in NI and the nominal part of the CPr, respectively). She, however, does not deal with scrambling, topicalization, and relativization of the preverbal noun in CPrs, and the separability of the CPr elements by these processes pose a challenge to the derivational framework she has adopted for analyzing these sequences. The incorporation analysis I put forward not only accounts for the bare nominals in Megerdoomian's analysis, but it can also be extended to include another type of multiword verbal expression in Persian, ADV+V sequences which are incorrectly treated in the literature as CPrs (Foli et al. 2005; Megerdoomian 2006; Pantcheva 2010).

In the next two sections, I will try to give an analysis of these two N+V sequences from the perspective of LFG. As my point of departure for distinguishing these two N+V sequences, I draw on the definition of CPrs given in Mohanan (1997) and Butt (1997) as a construction in which two semantically predicative elements jointly determine the argument structure of a single syntactic clause. The co-predication results in a complex argument structure, but a flat grammatical function structure, like that of a simple predicate. Based on this definition, cases of incorporation where an explicit or implicit argument of the verb and the verb make a complex are excluded, because incorporation does not give rise to a complex argument structure.

In order to account for cases of incorporation, I avail myself of the non-projecting words analysis put forward in Toivonen (2001), to argue that nouns as well as some adverbs in some ADV+V sequences are all non-projecting nodes, head adjoined to the V. As for the CPrs, I follow Butt (1995) and Alsina (1997) to account for the co-predication of their constitutive elements in terms of argument fusion.

3 Incorporation

That some of Persian multiword verbal expressions are the results of incorporation has already been discussed in Dabir-Moghadam (1997) and Vahedi Langrudi (1996). While I include their findings, in contrast to their conclusion, I claim that the results of incorporation are not CPrs. As observed in (2.2), these two N+V sequences contrast in their syntactic behaviour. NI resists separability by modification, relativization, scrambling and cliticization. These facts point to the lexical status of incorporation cases.

There are other facts discussed in Dabir-Mogaddam (1997), also in line with the list of defining characteristics in Mithun and Corbett (1999) that points towards the lexical status of Noun in the NI. Phonologically, they are pronounced as a whole, with no pause in between. They are treated as one

phonological word, with the noun bearing the primary stress. The pause after bare noun shows focus, and gives the noun prominence and a specific interpretation. They are highly productive and it can apply to most of the Object + Verb sequences to have an N+V incorporated sequence, as long as the noun is inanimate. Pragmatic factors, however, are at work to give a sequence the status of a well-established incorporation construct. *Gazā xordan* 'food-eating' is recognized by native speakers to function as a unitary activity and is packaged as conceptual whole by the native speaker. *havij xordan* 'carrot-eating' is not considered as such, although it has the potential to acquire the activity reading discerned in *Gazā xordan*. Semantically, they are transparent and the meaning of it corresponds to its parts. Even in the idiomatic ones, such transparency can be detected. Compare the two meanings of *dars xāndan* (Lit. 'lesson-reading') 'reading a lesson' vs. 'studying in an institute such as a university'.

Some other syntactic tests for constituency also illustrate that the Noun in NI is not a phrasal constituent. Due to the unavailability of the noun as a constituent on the c-structure, the following syntactic operations are not allowed: Gapping and the coordination of the incorporated noun with a specific noun (9), binding the pronominal (10), and nominal ellipsis (11).

(9) *man ham Gazā xord-am va ham mive rā
I also food eat-past.1.sg. and both fruit OM
Lit. '*I both food-ate and the fruit.'

(10) *man Gazā xord-am va kami az ān rā
I food eat-past.1.sg. and some from it OM
be gorbe dād-am
to cat give-past.1.sg.
Lit. '*I food-ate and gave some of it to the cat.'

(11) *ali Gazā xord va be bače-hā ham dād
Ali food eat.past.3.sg. and to child-Pl also give.past.3.sg.
Lit. '*Ali food-ate and gave (it) to the children.'

3.1 Is NI Lexical?

Mohanan (1995) introduces two different conceptions of lexicality: (a) 'lexical' as belonging to the lexicon as a module where items are formed; (b) lexical as the category of the unit formed. The data put forward so far illustrates that Persian NI can be lexical in both senses. There are, however, postlexical morphological facts that run counter to such an analysis. In the face of these facts, I argue that Persian NI is not created in the lexicon, but that it consists of a V^0 through head adjunction of the noun as a non-projecting lexical item with V^0 in the c-structure, as proposed in Toivonen (2001).

In many languages in the state-of-the-art on incorporation (Mohanan 1995, Mithun 1984, to name just two), the inflectional morphology appears on the edges of the N+V sequence as prefixes or suffixes, treating the whole sequence as a lexical item. In Persian, in contrast, the inflectional morphology appears on the verb as the host, thus intervening between the noun and the verb. These inflectional elements include the present tense (*mi-*), the subjunctive (*be-*), and the negative (*na-*) prefixes. If we assume that NI belongs to the lexicon, these facts are at odds with the assumptions of lexical morphology, based on which compounding (and incorporation as an instance of it) occurs before the word receives inflectional morphology. The prediction is that morpho-syntactic elements do not intervene between the elements, and they appear at the edges. This prediction is not attested in Persian NI as is illustrated by the following data.

(12) āryā **Gazā** *ne-mi-***xor-ad**
Arya food neg-IMP-eat-pres.3.sg.
'Arya doesn't eat food.'

It should be noted that *na-* in this construct has scope over the whole N+V and is not limited to the verb. To further clarify the facts about the scope of the negative marker, consider the following sentences. (The data is from spoken register.)

(13) āryā emruz hič **dars** **na-xund** (NI)
Arya today nothing lesson Neg-read.past.3.sg
'Arya didn't study at all today.'
(14) āryā emruz hič **dars**-iš ro **na-xund**
Arya today nothing lesson-3PossC OM Neg-read.past.3.sg
'Arya didn't study any of his lessons today.'

What (13) conveys is that Arya did not do the activity of reading (lit. lesson-reading) and the negative marker *na-* and the intensifying negative quantifier *hič* have scope over the whole activity denoted by N+V *dars xāndan* and consequently they have the whole sentence in their scope. In (14), the scope of *hič* is limited to the noun and *na-* has scope over the verb *xāndan* and as a consequence over the whole sentence. These facts about the scope of negative, then, point at the lexical status of the noun, while the separation by these prefixes questions the lexical status of the whole N+V sequence.

The future auxiliary *xāhad* also intervenes between the Noun and the Verb.

(15) āryā az aval-e mehr dar dāneSgāh-e tehrān
Arya from first-EZ Mehr in university-EZ Tehran
dars xāhad xānd
lesson FUT.3.sg. read
'Arya will begin his studies in the University of Tehran on 1st Mehr.'

Except for this functional word for future tense and affixes, Persian NI resists separation by content words and the noun remains adjacent to the verb.

3.2 Incorporated Nouns as Non-Projecting Words

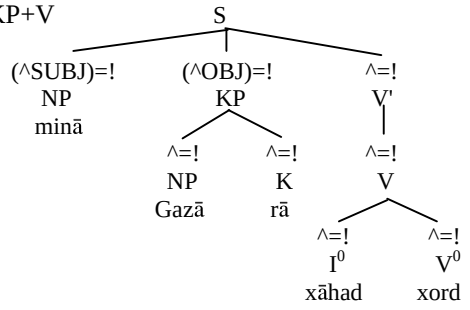
Considering the data in favor of the lexical status of the noun in NI and the data in (12) and (15), the analysis should consider the option of inflecting the verb for tense, aspect, person and number before they make a V^0. Persian verbs never appear in the root form at c-structure. They only appear as word forms, which is in line with the constraint of wordhood on the leafs in the c-structure, proposed by Bresnan (2001). The analysis that best captures the behavior of Persian NI is head-adjunction. In this analysis, the noun is treated as a non-projecting word that makes a V^0 when adjoined to a V^0 inflected for tense and aspect. Compare the following representations (18) and (19) for the NI (17) and its corresponding non-incorporated sentence (16).

To account for the close affinity of the future auxiliary and the other auxiliaries with the verb, head-adjunction of I^0 and V^0 is also posited, which is line with the recursive head adjunction discussed in Sadler (1998). Another explanation is due regarding the c-structure of Persian sentences, given its partial free word order. The Object (KP) and the Oblique (PP) receive case from the accusative marker *rā* and the preposition, respectively. Persian is also a pro-drop language, therefore the semantics of the Subject when dropped, is retrieved from the person/number agreement on the verb. Given that the grammatical functions in Persian are not dependent on their configurational positions in c-structure trees, a flat c-structure representation is posited for this language.

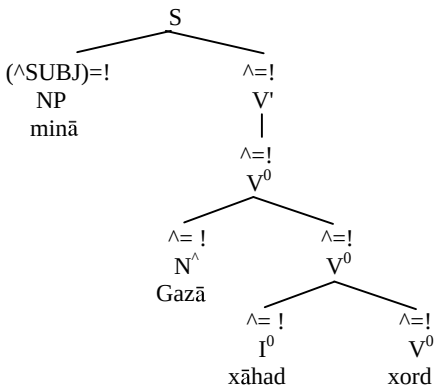
(16) minā Gazā rā xāhad xord
Mina food OM FUT.3.sg. eat-Past
'Mina will eat the food.'

(17) minā Gazā xāhad xord
Mina food FUT.3.sg. eat-Past
Lit. 'Mina food-ate.'

(18) c-structure for non-incorporated KP+V



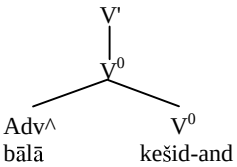
(19) c-structure for NI



As I have said above, the incorporation analysis can be extended to cover cases of the Adv+V sequences as well. As Persian is a free word order language with a flat c-structure, the verb can incorporate an adjacent argument or an adjunct, giving rise to new verbal complexes. These incorporated adverbs denote location and have the semantic role of Goal. As such they have the role predicted by Mithun to be among the roles that can be incorporated by a verb (Mithun 1984). Folli et al. (2005) have neglected this argument relation and classify these constructs as cases of CPr; but they correctly predict that Adv+V verbs have telic interpretations. These adverbs show lexical properties in that they resist modification and other constituency tests. In these cases, the Adv as a non-projecting word is head-adjoined to the V^0. As an instance of one such construction, consider (20) and the c-structure representation of the verbal complex in (21).

(20) xodešān u rā **bālā** **kešid-and**
themselves he OM up pull-3.Pl.Past
'They have promoted him themselves.'

(21)



The idiomaticity of the constructs obscures the argument relationship between the Adv and the V. Therefore, they have been treated as CPrs by most researchers (Karimi 1997; Folli et al 2005; Megerdoomian 2006; Pantcheva 2010). However, if one looks at the semantics of the verbal element in these sequences and the argument and adjunct positions it allows, and decomposes the idiomatic whole to its constitutive semantic components reflected in the literal meaning of each element of these sequences, the incorporation analysis seems plausible.

4 Complex Predication

CPrs in Persian consist of a nonverbal element and a light verb. The nonverbal element are claimed to be adjective, prepositional phrase, adverb or noun. As mentioned above, the scope of this study is limited to the noun and light verb combinations. In the section on the differences between NI and CPrs, we observed that the noun in CPrs is phrasal (it can be modified and it allows cliticization). It also allow scrambling and topicalization which are strong constituency tests. This fact about the phrasal status of the constitutive elements of CPrs is widely acknowledged in the state-of-the-art on Persian Cprs (Karimi 1997, Karmi-Doostan 1997, Megerdoomian 2006, Muller 2009). Karimi (1997) illustrates cases where the noun is modified by the quantifiers and adjectives. She also points at cases where the noun is separated by an intervening propositional phrase subcategorized by the noun in some of the CPrs (22).

(22) kimea unā ro **da'vat** be mehmuni **kard**
Kimea them rā invitation to party did
'Kimea invited them to a party.' (Karimi 1997: 281)

These cases cannot be straightforwardly accounted for in derivational approaches favored by most scholars in their studies of Persian CPrs: the noun must be accessible for both types of movements, that is the movements giving rise to CPr formation and also to undergo the syntactic operations of scrambling and modification (Vahedi-Langrudi 1996, Folli et al. 2005, Megerdoomian 2006, Pantcheva 2010). This has resulted in the fact that in the analyses, either only one type of movement has been the focus of the study or they have to assume many successive movements in the derivation.

LFG, as a theory which allows different independent interacting levels of representations, explains CPr formations by appealing to the predicate composition at c-structure and argument fusion at a-structure independently of each other. To account for complex predication in Persian, I draw on the theory of predicate composition proposed in Butt (1995) and Alsina (1997). In this theory, CPr formation occurs at the level of a-structure, which is independent of c-structure representations. The independence of c-structure allows for the possibility of discontinuous constituents mapping onto one single PRED value in the f-structure, a level of representation that is linked to both c-structure and a-structure.

In the remainder of this section, I deal first with CPr formation at the a-structure based on the argument fusion analysis developed in Butt (1995) and Alsina (1997). I avail myself of Pustejovsky's (1995) theory of Generative Lexicon towards a more fine-grained analysis of argument fusion and event fusion and also to be able to account for the cases where the noun combined with the light verb is not eventive. Then I attempt at c-structure analysis of Persian CPrs in line with Alsina (1997) in terms of a PRED sharing constraint.

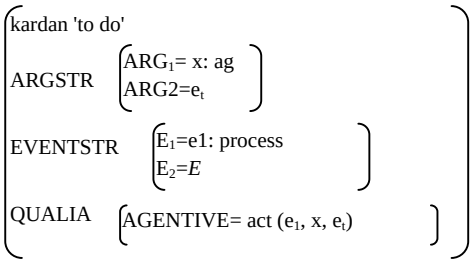
4.1 Argument Structure of Persian CPrs

According to the theory of predicate composition (Alsina 1997; Butt 1995), the argument structure of two semantic heads that carry the PRED value contribute to the overall argument structure or event structure of the CPr. What triggers argument fusion is an incomplete predicate which has a complete predicate in its argument structure. This complete predicate is represented as P* in Alsina (1997) to stand for any predicate that "will fully specify the underspecified argument structure of the incomplete predicate" (234). Butt (1995, 1997) states the same idea by integrating a Transparent Event ($\{E_T\}$) in the argument structure of the light verb. This $\{E_T\}$, which stands for an argument taking predicate, triggers CPr formation and the argument fusion of the highest argument of the embedded predicate with the lowest argument in the embedding incomplete predicate. I follow Butt's (1995) analysis of complex predicate formation and argument fusion with minor changes and also posit in the same line that the light verb is an incomplete predicate which selects for a transparent event, e_t. The eventive predicate combined with the light verb, then, maps onto this event, e_t. In the constructions under study, N+V sequences, the predicate that combines with the light verb is a nominal. Therefore this analysis should be modified to integrate mechanisms for mapping a noun onto an event regardless of the semantics of the noun as a predicative one as in *da'vat kardan* 'invitation doing' (to invite), or a non-predicative and non-eventive one as in *guš kardan* 'ear doing' (to listen). This calls for a deep lexical semantic theory of nominals as developed in Pustejovsky (1995).

Pustejovsky (1995) assumes a matrix for lexical semantic representation of words in terms of four different levels of argument structure (ARGSTR), event structure (EVENTSTR) and qualia structure (QUALIA) and lexical inheritance structure. Of these, the first three which are relevant for this study, will be posited in the representations. ARGSTR provides information about the number and the types of the arguments. EVENTSTR gives a description of an event in terms of its type (state, process, and transition), its internal structure and the subevents involved. QUALIA includes (a) the information of how an object and its constituents are related (CONSTITUTIVE role); (b) what distinguishes the object within a larger domain such as size, color (FORMAL role); (b) the purpose and the function of an object (TELIC role); and (d) factors involved in its origin or bringing it about (AGENTIVE role). Every word based on its type will be specified for the relevant kind of semantic information. What is activated is determined by the context in which the word is used. To clarify how this works in a

representation of light verbs, let us consider the following representation for *kardan* 'to do'[2].

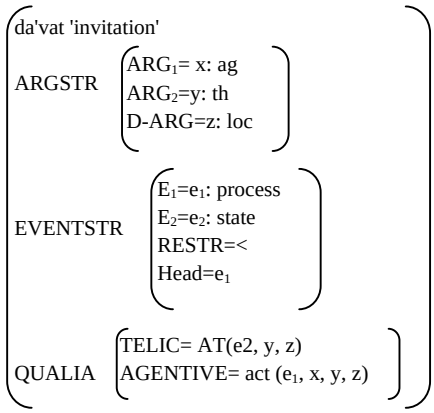
(23)



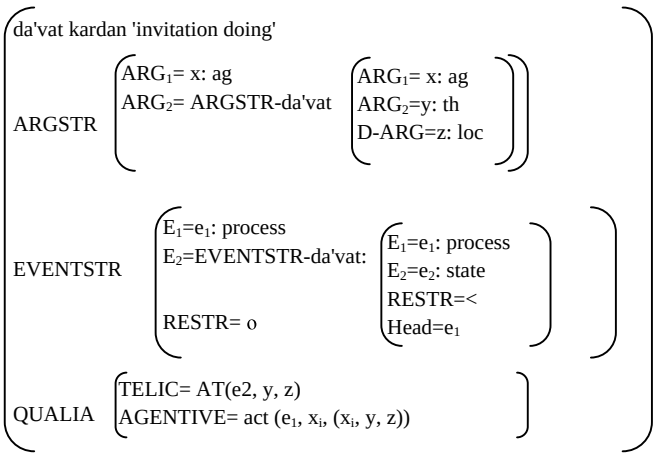
What this representation says is that the light verb *kardan* has two arguments: an x which is specified as the agent (I have departed from Pustejovsky's analysis in annotating the arguments for their thematic roles in order to serve the linking theory in mapping from arguments to grammatical functions), and an e_t which is the event which combines with the light verb to make it complete. In the event structure, it comprises of two events: e_1, a process which is the inherent event property of *kardan*, and the second one e_2 is the event contributed by the predicative noun and yet unspecified. The AGENTIVE role states that the agent (x) does the event denoted by e_2. When light verb combines with the eventive noun, the nominal's argument, event and qualia structure merge to give rise to a CPr. Following are the representations for *da'vat* 'invitation' (24) and the CPr formed as a result of the semantic information merging of these two predicates at different levels. (24) says that the eventive noun *da'vat* has three arguemnts, involved in an event made up of two subevents, process (of inviting) and state (of the invited being at z). RESTR in the EVENTSTR puts a restriction on the precedence of the events, here saying that the process subevent comes first. The two events are embodied in terms of TELIC role and AGENTIVE role in the QUALIA. The headedness principle says which subevent is the head in the internal structure of the event. The head subevent is the event that is linked to the syntax to be realized, which corresponds to TELIC role or AGENTIVE role. Here the head is process, which will be the one selected to fill in the e_t slot in the AGENTIVE role of *kardan*, as illustrated in (25).

2 Since there is not enough space for discussing the semantics of all the light verbs, the scope of the analysis will be further limited to the productive light verb *kardan* 'to do'. This light verb contributes agentivity to the complex.

(24)



(25)



The above representation illustrates how the different structures of these predicate merge to predicate jointly. The CPr inherits all the arguments of the constitutive predicates and in the level of the AGENTIVE role, the argument structure of the incomplete predicate is completed as the AGENTIVE role of the complete predicate is added. The presence of e_t triggers the merging of the two AGENTIVE roles. Accordingly, the higher argument of the embedded predicate is co-indexed with the only argument available for binding in the matrix predicate, since their thematic roles are compatible, both agents here. The composition of the two event structures into one follows from the event headedness principle: the event that is the head of the embedded predicate (process in *da'vat*) will be the event that is contributed at the level of event structure and maps to e_t in the AGENTIVE role of *kardan*. The composition ends in two simultaneous process events which leaves the question of headedness of the event at the matrix level irrelevant. The state event in the embedded event contributes to the *aktionsart* of the complex predicate, as reflected in the TELIC role of the CPr. Any event which has

process and state subevents in its internal structure, with process preceding the state and the process being the head subevent, is an accomplishment. *da'vat kardan* with this internal event structure is, therefore, an accomplishment.

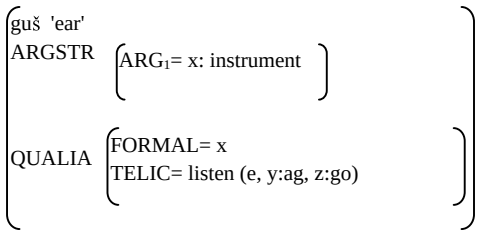
The predicative nouns have an event structure that satisfies the type required by the light verb to combine with, but the non-predicative nouns lack an event structure and do not have the event type required by the light verb. The combination of the light verb with a non-predicative noun results in a type clash, hence such a combination is predicted to be ruled out. The data, however, shows that it is possible for a non-predicative noun to combine with a light verb. *guš kardan* 'ear doing' (to listen) is one such construction, as used in (27). The assumption is that the QUALIA structures of the nouns have AGENTIVE and TELIC roles that make it possible for the light verb to select the event from there through the application of Selective Binding. Selective Binding is defined as follows (26).

(26) "SELECTIVE BINDING:
If α is of type <a,a>, β is of type b, and the qualia structure of β, QS_{β}, has quale, *q* of type a, the αβ is of type b, where $[[\alpha\beta]]=\beta\bigcap\alpha(q_{\beta})$." (Pustejovsky 1995: 129)

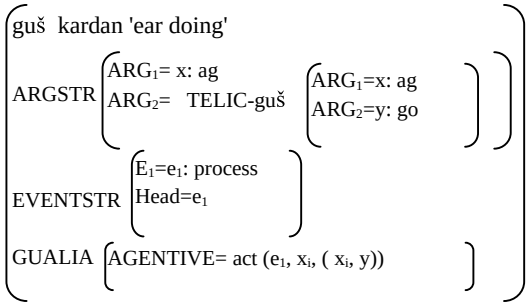
The application of the selective binding provides the light verb with the event type it requires to become complete. *kardan* looks into the representation for the non-predicative noun and it finds an event encoded in the TELIC role of the QUALIA of the noun, it selects that event and binds it into the ARGSTR. This event will be value of the e_t in the AGENTIVE role of *kardan.* The two representations below, for *guš* 'ear' (28) and the CPr *guš kardan* 'ear doing' (to listen) (29) are meant to clarify how selective binding works in the predicate composition of non-predicative *guš* with *kardan*. Notice that arguments represented for *guš* are those of the event denoted in its TELIC quale. Since the non-predicative noun does not have an event structure, it contributes nothing to the EVENTSTR. Thus, the event structure of this CPr has only a process in its internal event structure, giving rise to an activity reading. In the AGENTIVE role, the two events merge and the identical arguments of them are co-indexed and unified.

(27) man diruz be rādiyo guš kardam
I yesterday to radio ear do-Past.1.sg.
'Yesterday, I listened to the radio.'

(28)



(29)



In order to map the arguments onto the relevant grammatical functions, lexical mapping theory (LMT), as discussed in Butt (1995) and Bresnan (2001), applies. Before applying LMT, the headedness principle (Pustejovsky 1995) determines which arguments are mapped obligatory and which arguments are optional. The arguments in the head event are obligatorily mapped onto the grammatical functions. The f-structure representations as a result of mapping from arguments onto grammatical functions for the verbs *da'vat kardan* (30) and *guš kardan* (31) are given below.

(30) da'vat kardan: AGENTIVE= act (e_1, x_i, (x_j, y, z))

ag	th	loc
-o	-r	-o
-r		+r
SUBJ	OBJ	OBL

(31) guš kardan: AGENTIVE= act (e_1, x_i, (x_i, y))

ag	go
-o	-o
-r	+r
SUBJ	OBL

As it can be read from the above representation, the two predicates are combined to make one predicate on the a-structure and f-structure levels and they share one argument structure.

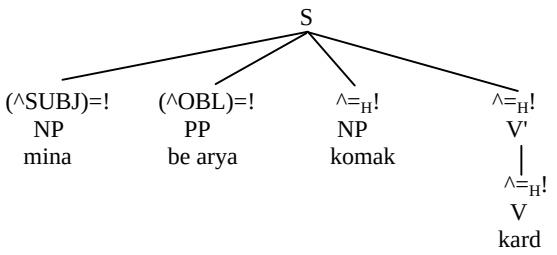
4.2 Tree Structure of Persian CPrs

As concluded above, Persian CPrs map on one PRED value at the level of f-structure. The constitutive elements of CPrs, however, are not adjacent in the c-structure and the noun shows properties typically associated with phrasal categories, such as modification, relativization and scrambling. In order to allow for two constituents specified with the PRED value to compose on the c-structure and at the same time to account for discontinuous constituents of CPrs on the phrasal structure level, Alsina (1997) suggests annotating the relevant nodes of the constitutive elements with $\uparrow=_H!$. This constraint says that f-structure values of the mother except for the PRED value are unified and the PRED value of the mother node is the result of the composition of the PRED value of that node with that of its head sister constituents. To illustrate how this works, c-structure representations for sentences (32) and (33) in are given in (34) and (35), respectively.

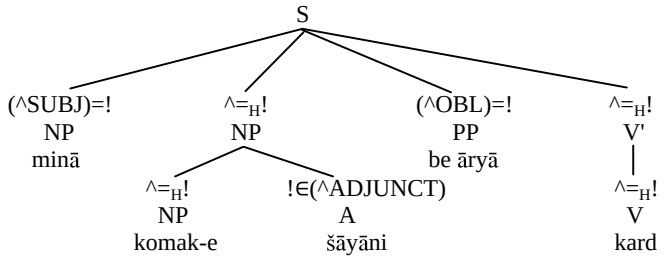
(32) minā be āryā komak kard
Mina to Arya help do.past.3.sg
'Mina helped Arya.'

(33) minā komak-e šāyāni be āryā kard
Mina help -EZ considerable to Arya do.past.3.sg
'Mina helped Arya a lot.'

(34)



(35)



The c-structure (34) shows that both the light verb and the noun are annotated by a PRED composition constraint, $\uparrow=_H!$, requiring the PRED value of the annotated nodes to be composed to the PRED value of the

mother node. The tree structure in (35) also illustrates that the scope of modification is not limited to the noun *komak* 'help' and the PRED value is shared by the whole CPr.

5 Conclusion

N+V sequences show different syntactic behaviors. Some of these N+V sequences are classified as NI because there is an argument relationship between the noun and the verb and there is a strict constraint on the adjacency of the noun and the verb. NI cases are analyzed based on head-adjunction of non-projecting words proposed by Toivonen (2001). Based on this analysis, the noun has a lexical status and is not able to project as a phrasal category. The other N+V sequences are discussed as cases of complex predication, where the noun combines with an incomplete predicate to project on a single PRED value with a shared argument structure. To analyze CPrs, I used the CPr formation analysis proposed in Butt (1995, 1997) and Alsina (1997) along with adaptations from Pustejovsky's multi-layered semantic representations. To extend the analysis of the Noun+V CPrs from eventive nouns to non-eventive nouns, selective binding (Pustejovsky 1995) of semantic information was considered. Finally, an attempt is made to account for the discontinuous constituent structure of CPrs by employing the PRED composition constraint developed in Alsina(1997) to apply on the c-structure representations. The scope of this research was limited to N+V sequences and in the analysis of CPrs I limited the scope to the light verb *kardan* 'to do'.

References

Alsina, Alex. 1997. A Theory of Complex Predicates: Evidence from Causatives in Bantu and Romance. In Alex Alsina, et al. (eds.), *Complex Predicates*, pages 203-246 . Stanford: CSLI Publications.

Bresnan, Joan. 2001. *Lexical Functional Syntax*. Oxford: Blackwell Publishing.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications.

Butt, Miriam. 1997. Complex Predicates in Urdu. In Alex Alsina, et al. (eds.), *Complex Predicates*, pages 107-149. Stanford: CSLI Publications.

Dabir-Moghaddam, Mohammad. 1997. Compound Verbs in Persian. *Studies in Linguistic Sciences* 27, 25-59.

Folli, Raffaella, Heidi Harley, and Simin Karimi. 2005. Determinants of Event Type in Persian Complex Predicates. *Lingua* 115(10),1365–1401.

Karimi, Simin (1997) Persian Complex Verbs: Idiomatic or Compositional? *Lexicology* 3.1, 273-318.

Karimi-Doostan, Gholamhossein. 1997. *Light Verbs Constructions in Persian*. Ph.D. thesis, University of Essex.

Megerdoomian, Karine. 2006. The Status of the Nominal in Persian Complex Predicates. available at http://www.zoorna.org/papers/preverbs.pdf

Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60, 847–894.

Mithun, Marriane and Gerville Corbett. 1999. The effect of noun incorporation on argument structure. In Lunella Mereu (ed) *The Boundaries of Morphology and Syntax*, pages 49-71. Amsterdam: John Benjamins.

Mohanan, Tara. 1995. Wordhood and Lexicality: Noun Incorporation in Hindi. *Natural Language and Linguistic Theory* 13, 75-134.

Mohanan, Tara. 1997. Multidimensionality of Representation: NV Complex Predicates in Hindi. In Alex Alsina, et al. (eds.), *Complex Predicates*, pages 431-471. Stanford: CSLI Publications.

Pantcheva, Marina. 2010. First Phase Syntax of Persian Complex Predicates: Argument Structure and Telicity. *Journal of South Asian Linguistics* 2, 53-72.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge: MIT Press.

Sadler, Louisa. 1998. Welsh NPs without Head Movement. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG98 Conference*, CSLI On-line Publications.

Toivonen, Ida. 2001. *The Phrase Structure of Non-Projecting Words*. Ph.D. Thesis, Standford University.

Vahedi-Langrudi, Mohammad-Mehdi. 1996. *The Syntax, Semantics and Argument Structure of Complex Predicates in Modern Farsi*. Ph.D. thesis, University of Ottawa.