

Restrictive Relative Clauses in Persian: An HPSG Analysis

Mehran A Taghvaipour
University of Essex
Colchester, CO4 3SQ, UK
matagh@essex.ac.uk

1 Introduction

This paper presents an analysis for Persian restrictive relative clauses (RCs) in the Head-driven Phrase Structure Grammar (HPSG) framework. I will first provide an overview of some general properties of the language, with a particular emphasis on RCs, resumptive pronouns (RPs), their pattern of distribution, and their similarity with gaps in RC constructions. In Section 3, I will review two relevant HPSG works on Unbounded Dependency Constructions (UDCs): Pollard and Sag (1994) and Sag (1997). Section 4 presents my own analysis. At the bottom of the dependency, I will assume a special *sign* that has a nonempty value for the SLASH feature. This special sign can be either a RP or a trace. I will introduce a feature called GAPTYPE which is a non-local feature and whose value can be either *trace* or *rp*. I will introduce two constraints to capture the pattern of distribution of RPs and traces. In the middle of the dependency, a trace or a RP is treated entirely in terms of the inheritance of the SLASH feature as in Sag (1997). At the top of the dependency, I will stop the non-empty SLASH at the complementizer point. I will propose a lexical entry for the complementizer *ke* that will account for the binding of SLASH by the feature BIND, which has a non-empty set as value. Section 5 summarises the paper.

2 The Persian Language

Persian is a null-subject verb final language with SOV word order. Example (1) represents a simple sentence in Persian.

- (1)
(*mæn*) *ye doxtær daræm.*
(I) a daughter have-PRES-1sg
'I have a daughter.'

RCs in Persian, as in English, are head-modifying constituents in the sense that they modify the NP they follow. A typical Persian RC is always introduced by the invariant complementizer *ke*. Example (2) shows a restrictive RC in brackets.

- (2)
zæn-i¹ [ke mæn dust+daræm]
woman-RES COMP I like-PRES-1sg
'The woman that I like'

What follows in (3) is a schematic representation of a Persian nominal containing a RC.

- (3)
-
- ```
graph TD
 N_prime[N'] --- N_prime_left[N']
 N_prime --- RC[RC]
 RC --- ke[ke]
 RC --- S[S]
```

The grammar of Persian allows personal pronouns to behave resumptively. That is, a personal pronoun is allowed to be used where a gap might be expected. RCs are one of the Persian constructions in which RPs are used. A RP comes inside a RC and is co-indexed with the noun modified by the RC. Example (4) represents a Persian RC in which the pronoun *u*, i.e. 's/he', is used resumptively.

- (4)  
*mærd-i ke be u pul dadim*  
man-RES COMP to **him** money gave-1pl  
'the man to whom we gave money'

---

<sup>1</sup> This particle precedes restrictive RCs in Persian and is shown, henceforth, by RES in gloss.

Table 1 below shows the pattern of distribution of gaps and resumptive pronouns in Persian restrictive RCs.

|              | Subject | Object of Prep. | Genitive | Direct Object |
|--------------|---------|-----------------|----------|---------------|
| Gap Allowed? | Yes     | No              | No       | Yes           |
| RP Allowed   | No      | Yes             | Yes      | Yes           |

TABLE 1: GAPS & RPS IN PERSIAN RESTRICTIVE RCs

Table 1 shows that in some positions, only gaps are allowed. In other positions only RPs are allowed. Both gaps and RPs are alternatively possible in some other positions.

As illustrated in Table 1, if the relativized position is subject, a RP cannot appear. Examples in (5) illustrate this.

(5a)

*mærd-i ke \_\_\_ pirahæn-o pušid*  
 man-RES COMP \_\_\_ shirt-RA wore-3sg  
 ‘The man who wore a shirt...’

(5b)

*\*mærd-i ke u pirahæn-o pušid*  
 man-RES COMP **he** shirt-RA wore-3sg  
 ‘The man who wore a shirt...’

If the position relativized is object of preposition, the presence of a RP is obligatory; otherwise the result is ungrammatical as shown in (6b).

(6a)

*mærd-i ke æz u pul gerefti*  
 man-RES COMP from **him** money took-2sg  
 ‘The man from whom you took money ...’

(6b)

*\*mærd-i ke æz \_\_\_ pul gerefti*  
 man-RES COMP from  $\emptyset$  money took-2sg  
 ‘The man from whom you took money ...’

Table 1 also shows that if the position relativized is the possessor position, a RP must be present. This is contrasted in (7a) and (7b).

(7a)

*mærd-i ke pirahæn-e<sup>2</sup> u zærd æst*  
 man-RES COMP shirt-EZ **he** yellow is  
 ‘The man whose shirt is yellow ...’

(7b)

*\*mærd-i ke pirahæn-e \_\_\_ zærd æst*  
 man-RES COMP shirt-EZ  $\emptyset$  yellow is  
 ‘The man whose shirt is yellow ...’

Table 1 illustrates how Persian allows gaps and RPs alternatively if the relativized position is direct object. Example (8), taken from Safavi (1994: 187) shows this possibility as both readings are grammatical. This alternative possibility argues against the idea that RPs are a ‘last resort’ device (Shlonsky, 1992).

(8)

*hušæng ketab-i ra<sup>3</sup> ke pesaræm*  
 Hushang book-RES RA COMP son-my

*(an ra) xarideh-bud dozdid*  
 (it RA) buy-PP-3sg steal-PAST-3sg

‘Hushang stole the book that my son had bought for me.’

Later on in Section 4 where I provide my HPSG analysis, I will argue against the assumption that the features responsible for gaps and RPs have different values. As my analysis is based on the assumption that they both have the same type of value, I will provide below a variety of evidence in favour of the similarity of gaps and RPs from the following phenomena: (i) coordinate structures, (ii) parasitic gaps, and (iii) island constraint.

A strong argument in support of the fundamental similarity of RPs and gaps are comes from

<sup>2</sup> *Ezafæ* particle. This particle is pronounced /e/ and is shown, henceforth, by –EZ in gloss. This particle structurally links nouns with their modifiers or possessors.

<sup>3</sup> This particle is a marker for specificity.

coordinate structures. Example (9) shows how in Persian a RP can be used with a gap in coordinate structures in unbounded dependencies. In fact, in this language, it is possible to have gaps in both conjuncts, RPs in both, or a gap in one conjunct and a RP in the other (in any order).

(9)  
*mærd-i ke \_\_\_\_\_ pirahæn-e zærd*  
 man-RES COMP \_\_\_\_\_ shirt-EZ yellow

*pušideh+bud væ shoma*  
 wear-PRESPART-3sg and you

*diruz az u pul*  
 yesterday from **him** money

*qærz+gereftid Ali bud.*  
 borrow-PAST-2pl Ali was

‘The man who \_\_\_\_\_ was wearing a yellow shirt and you borrowed money from was Ali.’

Another argument that supports the similarity between RPs and gaps comes from parasitic gaps. Persian data shows that RPs, like gaps, can license parasitic gaps. I shall bring examples (10a) and (10b) to illustrate this possibility. In (10a) there are two gaps, the second of which is parasitic. (10b) shows a sentence in which the second gap is still parasitic but it is licensed by the RP *un*.

(10a)  
*in ketab-i-ye ke Yasmin*  
 this book-RES-is COMP Yasmin

*bedun in ke \_\_\_\_\_ bexuneh*  
 without this COMP \_\_\_\_\_ read-3sg

\_\_\_\_\_ *xærid.*  
 Ø bought-3sg.

‘This is the book that Yasmin bought \_\_\_\_\_ without reading \_\_\_\_\_’

(10b)  
*in ketab-i-ye ke Yasmin*  
 this book-RES-is COMP Yasmin

*bedun in ke un ro bexuneh*  
 without this COMP **it** RA read-3sg

\_\_\_\_\_ *xærid.*  
 Ø bought-3sg.

‘This is the book that Yasmin bought (**it**) without reading \_\_\_\_\_’

Another piece of supporting evidence for the similarity of RPs and gaps in Persian is the sensitivity of RPs, like gaps, to certain islands. For instance, Persian gaps are sensitive to Subject Condition Constraint as illustrated in the pair of sentences in (11).

(11a)  
*[in ede’a ke Ali Hæmid ra dideh]*  
 [this claim COMP Ali Hamid RA seen

*Yasmin ra narahat kærd.*  
 Yasmin RA annoyed

‘The claim that Ali has seen Hamid annoyed Yasmin.’

(11b)  
 \**mærd-i ra ke [in ede’a ke Ali \_\_\_\_\_/*  
 man-RES RA COMP [this claim COMP Ali \_\_\_\_\_/

*u ra dideh] Yasmin ra narahat kærd.*  
 him RA seen] Yasmin RA annoyed.

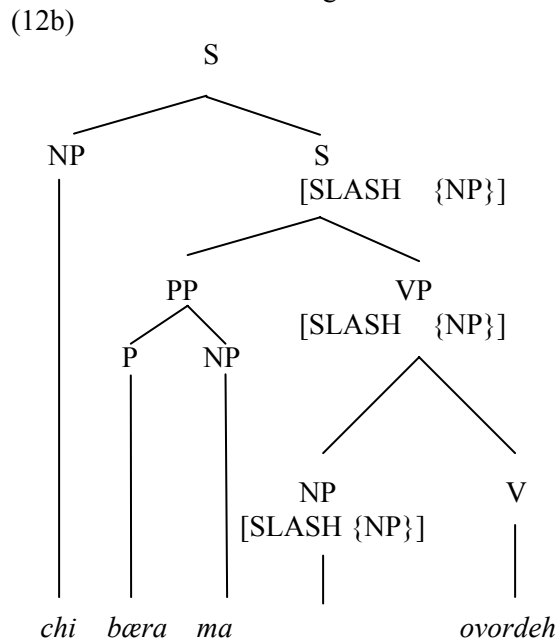
‘The man that the claim that Ali has seen \_\_\_\_\_/him annoyed Yasmin.’

### 3 UDCs in HPSG

RCs in Persian are instances of Unbounded Dependency Constructions (UDCs). They always contain either a gap or a RP, which is linked to and licensed by the noun modified by the RC.

Most HPSG approaches to UDCs involve a feature called SLASH. This feature encodes the fact that a constituent contains an unbounded dependency gap of some sort. SLASH appears within every category between the gap and the licensing higher structure as shown in (12b).

(12a) *Chi bæra ma ovorde?*  
 What for us bring-PRESP-3sg  
 ‘What has he brought for us?’

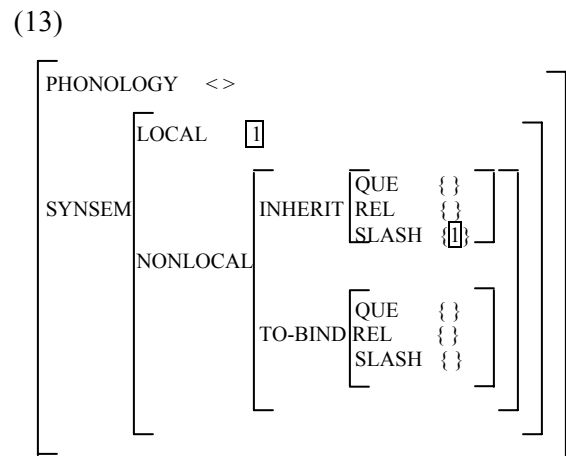


Structures like the one in (12) above contain three distinct parts:

1. The bottom: where there is the gap of the form of [SLASH {NP}];
2. The middle: consisting of local trees with a SLASH feature on both daughter and its mother;
3. The top: in which a SLASH feature appears on a daughter but not on the mother.

Pollard and Sag (1994: 161) assume that every unbounded dependency is introduced at a terminal node (at the bottom of the dependency) by a special sign that has a nonempty value for the appropriate nonlocal feature. In relative constructions in English, this sign is a relative *wh*-word (e.g. *who*, *which*), containing a nonempty value for the REL feature. In *wh*-questions, it is an

interrogative *wh*-word with a nonempty value for the QUE feature. And, in filler-gap dependencies, this special sign is a *trace* with a nonempty value for the SLASH feature. They give the lexical entry in (13) for *trace*, a special lexical item that has no phonology.



As shown in (13), the only nonempty value in nonlocal features of a trace is the SLASH value, which is a singleton set. As shown by tag [1], the member of this set is just whatever the local features of the trace are. Once a trace occurs as a complement of some head, then it will structure-share whatever local features are specified for that complement by the head. In turn, whatever those local features are, they will show up in the SLASH value set of the trace.

Pollard and Sag (1994) use an INHERIT feature to handle inheritance of SLASH in the middle of the dependency. The information that there is an unbound trace is encoded in SLASH and is propagated up the tree. The mechanism responsible for this propagation is one of the principles of universal grammar posited in HPSG, called the Nonlocal Feature Principle (NFP) as given in (14).

(14) NONLOCAL FEATURE PRINCIPLE

For each nonlocal feature, the INHERITED value on the mother is the union of the INHERITED values on the daughters minus the TO-BIND value on the head daughter.



Amalgamation of SLASH and is given in (18) below.

(18) Lexical Amalgamation of SLASH

$$word \Rightarrow \left[ \begin{array}{l} \text{BIND } \boxed{0} \\ \text{ARG-ST } \langle [\text{SLASH } \boxed{1}], \dots, [\text{SLASH } \boxed{n}] \rangle \\ \text{SLASH } (\boxed{1} + \dots + \boxed{n}) - \boxed{0} \end{array} \right]$$

All words, except SLASH binding elements like *tough*, specify empty value for the feature BIND. That is, in most cases nothing is subtracted from the disjoint union of the argument's SLASH values. Therefore, if a non-head-daughter is slashed so should the head daughter. For instance, if a verb's complement is slashed, the verb itself is slashed.

In Sag (1997), the SLASH is inherited upward in terms of the SLASH Inheritance Principle (SLIP) in (19).

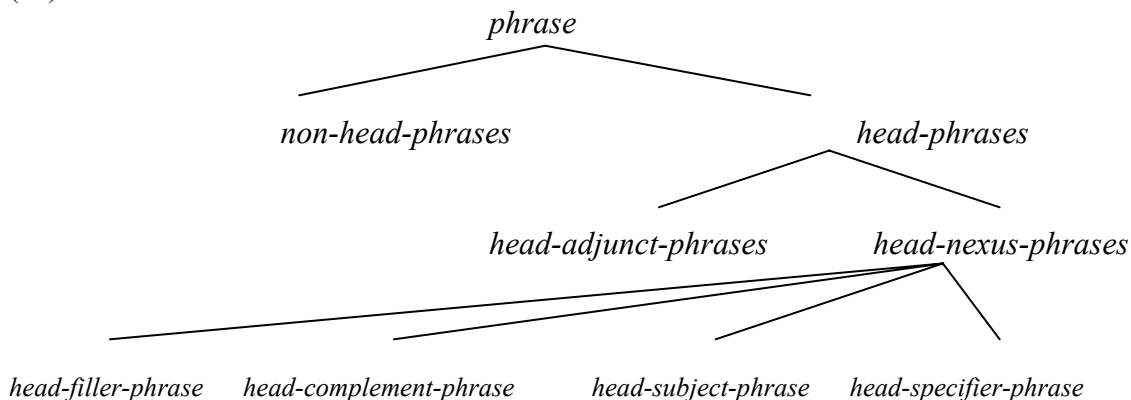
(19) SLASH Inheritance Principle (SLIP):

$$hd\text{-nexus-ph} \Rightarrow \left[ \begin{array}{l} \text{SLASH } / \boxed{1} \\ \text{HD-DTR } / [\text{SLASH } \boxed{1}] \end{array} \right]$$

This constraint guarantees that the SLASH value of a phrase (of the type head-nexus-phrase) is the SLASH value of its head-daughter. In this way, any SLASH inheritance is mediated by the head-daughter, whose SLASH value contains that of the relevant non-head daughter.

In Sag (1997), the binding off of the SLASH specification, at the top of the dependency, occurs at an appropriate point higher in the structure through various constraints associated with the various clausal subtypes. Sag (1997) assumes that phrases in English and other languages can be classified according to the hierarchy shown in (20).

(20)

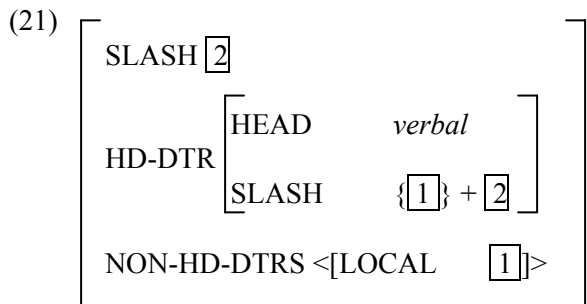


Sag (1997) further suggests multidimensional classification of phrases by introducing a CLAUSALITY dimension. This dimension distinguishes clauses from non-clauses, and subcategorises clauses into four subtypes: *imperative clauses*, *declarative clauses*, *interrogative clauses* and *relative clauses*. For *wh*-relative clauses, Sag (1997) suggests the two dimensional hierarchy. From one dimension, these

relative clauses inherit properties and constraints from *head-nexus-phrases* (*hd-nexus-ph*) and its subtypes. From the other dimension, they inherit properties and constraint from *relative-clause* type and its subtypes.

Returning to our topic in this section, I will examine the binding off of SLASH in a RC like *who they like*. Being a *fin-wh-fill-rel-cl* (finite-*wh*-

filler-relative-clause), this clause is subject to the constraint on *hd-fill-ph* types, as given in (21).



According to this constraint, the binding off of the head daughter's SLASH value is guaranteed by index sharing between the LOCAL value of the NON-HD-DTRS and the SLASH value of the HD-DTR. In other words, this constraint ensures that LOCAL value of the non-head daughter is identical to one member of the SLASH set and that the SLASH set of the phrase is identical that of the head minus this member.

#### 4 An HPSG Analysis for Persian RCs

In Section 2 above, I noted that Persian RCs, like their English counterparts, follow the NP they modify. I also illustrated that Persian RCs, unlike their English counterparts, do not contain any relative pronoun; instead, they all have the invariant complementizer *ke*. Furthermore, I demonstrated that Persian RCs all contain either a gap or a RP. The pattern of distribution of gaps and RPs is such that there are some positions where only gaps are allowed; there are other positions where only RPs are allowed; and yet there are also some other positions where both gaps and RPs are alternatively allowed. Finally, I showed that gaps and RPs in Persian RCs are strikingly similar: they can be coordinated with each other in a conjunct; they can both license parasitic gaps; and, they are both sensitive to some island constraints.

As there is no relative pronoun involved in Persian RC constructions, the analysis of Pollard and Sag (1994) for the top of the dependency does not

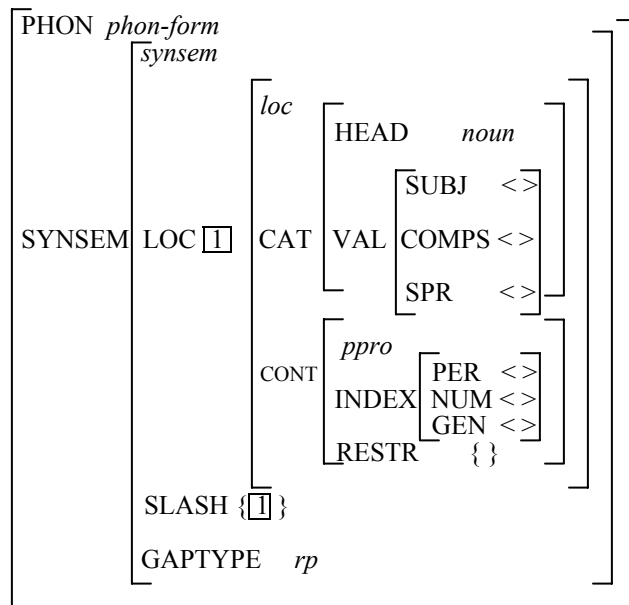
apply to Persian RCs as there is no filler-gap dependency involved.

In addition, as all RCs in Persian contain the complementizer *ke*, it is unnecessary to develop a complex classification like the one developed in Sag (1997). Given that all Persian RCs share an overt lexical element (i.e., *ke*), I will attribute the properties of Persian RCs mainly to this element.

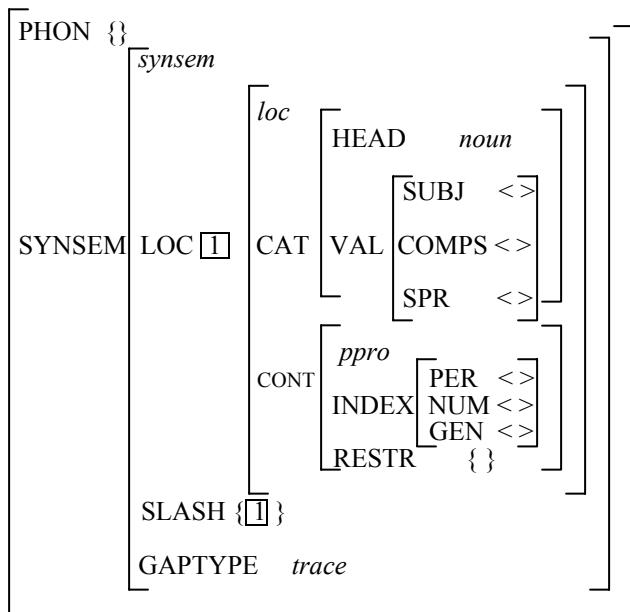
I shall assume that the unbounded dependency in Persian RCs appear at the bottom of the dependency by a special sign that has a nonempty value for the SLASH feature. This special sign is either a trace or a RP. This nonempty SLASH feature encodes the information that there is a dependency between the gap/RP and the NP modified by the RC.

Due to the fact that RPs and gaps are strikingly similar in Persian RCs, I shall assume that Persian RCs contain a trace where RPs are not present. I shall propose the lexical entry in (22) for RPs and the one in (23) for traces.

#### (22) Lexical Entry for a resumptive pronoun



(23) Lexical Entry for a *trace*



The lexical entries in (22) and (23) are the same except in two respects. Firstly, the value of the PHON feature in traces is an empty set. This means that RPs as overt elements have phonology but traces do not.

The second difference between these two lexical entries is that the value of their GAPTYPE features is different. GAPTYPE is a feature that I have introduced in order to capture the distributional properties of RPs and traces. GAPTYPE is a non-local feature whose value can be either *trace* or *rp*, for traces and RPs respectively.

As for the pattern of distribution of RPs and traces, we, first need to prevent RPs from appearing in subject position. I will propose the constraint in (24) to deal with this.

$$(24) \quad [\text{SUBJ } <[1] >] \rightarrow \sim ([1] = [\text{SYNSEM}| \text{NONLOC}| \text{GAPTYPE } rp ])$$

The effect of this constraint is that if an element is in subject position, then the value of its GAPTYPE feature cannot be *rp*. In other words, if an element is a RP whose value of the GAPTYPE

feature is *rp*, then it cannot come in subject position.

The second constraint, we will need here is one to prevent traces from appearing in the positions of object of prepositions and possessors (i.e., in positions of the complements of non-verbs). This constraint is proposed in (25).

$$(25) \quad \left[ \begin{array}{l} \text{HEAD } [1] \\ \text{COMPS } < \dots, [\text{GAPTYPE } trace], \dots > \end{array} \right] \rightarrow [1] = \text{verb}$$

The effect of this constraint is that if there is a trace as a complement of a head, then that head has to be a verb. Therefore, as in the case of object of preposition and genitive cases (possessors), the head is not a verb, we will not have a trace therein.

In the middle of the dependency, I will follow Sag (1997). The SLASH is inherited upward by two constraints, given in (18) and (19) above.

At the top of the dependency, we will need some way to bind the SLASH feature. In other words, we will need a way to ensure that the non-empty SLASH value stops at an appropriate point. This appropriate point, in Persian RCs, is the complementizer *ke*. I shall propose the lexical entry in (26) for *ke* in RCs (i.e., *ke<sub>RC</sub>*).

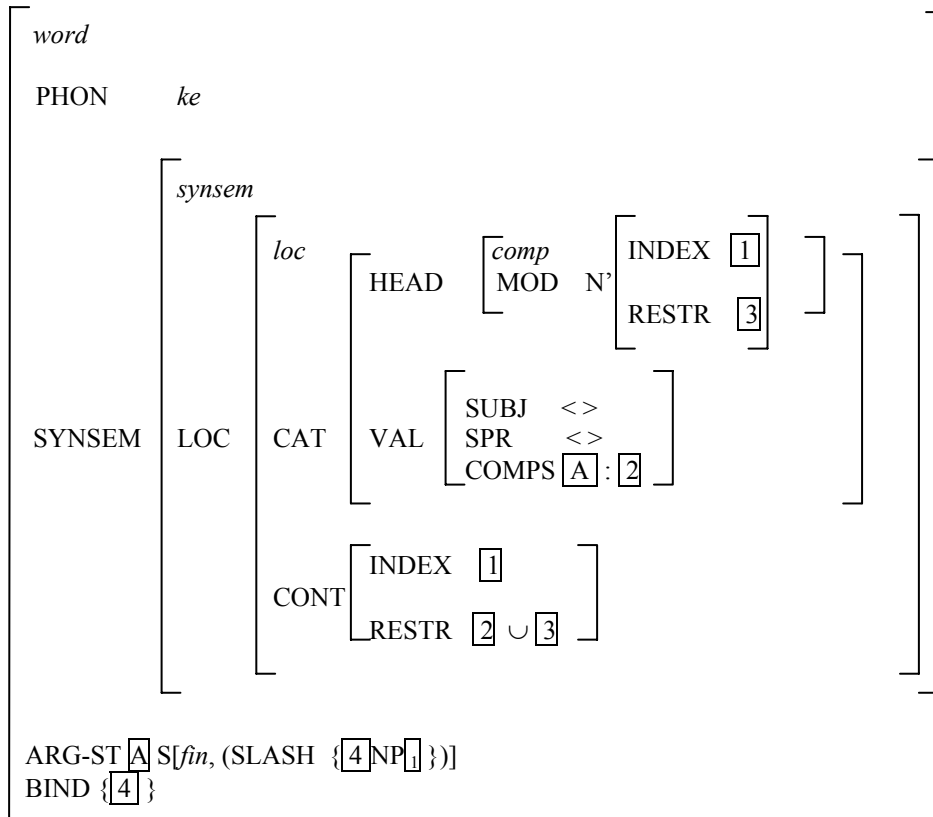
The lexical entry for *ke* specifies some lexical information that ensures that the index of the N' (the NP modified by the RC) is identical to the SLASH value of *ke*. This structure-sharing, which is shown by tag [1], relates the gap or the resumptive pronoun to the NP modified by the RC.

In addition, (26) ensures that *ke* requires a sentential complement, shown by tag [A]. Tag [A] is the only member of *ke*'s ARG-ST list that stands for a finite sentence, containing a trace or a RP.

The lexical binding of SLASH is accounted for by the feature BIND, which has a non-empty set as value for *ke*. This is shown by tag [4]. The BIND

feature will ensure that the trace or the RP is not amalgamated into the SLASH value of *ke* itself.

(26) Lexical Entry for *ke<sub>RC</sub>*



**5 Summary**

This paper has presented an analysis for Persian restrictive RCs in HPSG. After providing an overview of some general properties of Persian RCs, resumptive pronouns (RPs), and their pattern of distribution, as well as their striking similarity with gaps, I reviewed two relevant HPSG works: Pollard and Sag (1994) and Sag (1997). In Section 4, I presented my own analysis. I showed how the dependency in RCs is encoded in the structure by the SLASH feature. I showed how this dependency appears at the bottom of the dependency, how it is inherited from daughters to mothers via head-daughter, and how it is terminated by the non-empty BIND feature of the complementizer *ke*.

**References**

- Pollard C., and I. Sag, 1994. Head-Driven Phrase Structure Grammar. The University of Chicago Press, USA.
- Safavi, K., 1994. Some Properties of Persian Relative Clauses. Proceedings of the 2<sup>nd</sup> Conference on Persian Linguistics, Allameh Tabatabaie University, Tehran.
- Sag, I., 1997. English Relative Clause Constructions. Journal of Linguistics 33:431-484.
- Shlonsky, U., 1992. Resumptive Pronouns as a Last Resort. Linguistics Inquiry 23/3: 443-468.