

# **A Computational Analysis of the Persian Noun Phrase**

*Karine Megerdooian*

Memoranda in Computer and Cognitive Science  
MCCS-00-321

Computing Research Laboratory  
New Mexico State University  
Las Cruces, New Mexico  
April 2000

---

## **Abstract**

The highly ambiguous structure of the Persian Noun Phrase (NP) causes immense difficulties for automatic parsing of written text. Factors contributing to the ambiguity of the Persian NP structure include the lack of overt morphology to mark boundaries, the fact that short vowels are not written in Persian text, a relatively free word order and the optionality of the subject. This paper introduces the constituents forming a Noun Phrase in Persian. The syntax of NPs and relative clauses is described. A closer study of the Noun Phrase provides lexical and morphological clues for determining boundaries. To describe the NP rules in the Shiraz project, a unification-based syntactic grammar was used, which operates on typed feature structures.

---

# Contents

<b>1. Introduction.....</b>	<b>1</b>
<b>2. Basic Noun Phrase .....</b>	<b>2</b>
2.1 Constituents .....	2
2.2 Simple Noun Phrase.....	3
2.2.1 Standard NP .....	3
2.2.2 Infinitival Noun Phrase .....	4
2.2.3 Possessor Noun Phrase .....	4
2.3 Complex Noun Phrase .....	5
<b>3. Noun Phrase Boundaries.....</b>	<b>6</b>
3.1 Lexical Categories .....	6
3.2 Indefinite Article / Enclitic Particle (IE).....	7
3.3 Personal Pronoun Clitics.....	7
3.4 Ezafe Morpheme.....	8
3.5 Combination of Boundary-Marking Features.....	8
<b>4. Bolero Grammar.....</b>	<b>10</b>
<b>5. Relative Clause.....</b>	<b>12</b>
<b>6. Conclusion .....</b>	<b>14</b>
<b>Appendix.....</b>	<b>15</b>

# 1

---

## Introduction

The highly ambiguous structure of the Persian Noun Phrase (NP) causes immense difficulties for automatic parsing of written text. Numerous factors contribute to the ambiguity of the Persian NP structure. Certain vowels, known as short vowels, are not written, which produces lexical ambiguities. There are very few overt morphemes in the language to mark boundaries of Noun Phrases: With the exception of the specific object marker, the language lacks Case morphology. There are often no particles in written text linking the constituents of a Noun Phrase, such as “of” in English, since these particles are pronounced as short vowels and are therefore not transcribed in written form. Furthermore, since the basic word order in Persian is Subject-Object-Verb, the lack of overt morphology for marking boundaries makes it very difficult to determine where the Subject ends and the Object begins. All of these factors, coupled with a relatively free word order and the optionality of the subject, combine to make the Persian Noun Phrase extremely ambiguous for an analysis of written text.

This paper introduces the constituents forming a Noun Phrase in Persian, as well as a description of its structure. This report shows how the lexical and morphological information present in the written text could be used in determining the boundaries of the NP. To describe the NP rules in the Shiraz project, a unification-based syntactic grammar was used. This grammar, known as Bolero, operates on typed feature structures. Section 4 discusses several of these rules. More complex constructions, such as Relative Clauses, are discussed in the final section of this report. The romanization used in transcribing the Persian examples is given in the Appendix.

## 2

---

### Basic Noun Phrase

The structure of the basic Noun Phrase in Persian is discussed in this section. Complex constructions such as the Relative Clause are described in Section 5. We distinguish between the *simple* and the *complex* Noun Phrase. A complex NP is the noun phrase structure that is formed from several simple NPs. In this section, we will describe the simple NP constructions and in Section 2.3, we will discuss the complex NP structure.

#### 2.1 Constituents

The main constituents of a Noun Phrase in Persian are introduced below. These parts of speech will be used in later sections when discussing the structure of the NP. Note that there is no definite article in Persian, only an indefinite, which appears as an affix attached to the noun or adjective.

**Noun.** The Noun is the head of the Noun Phrase.

**Adjective.** Adjectives modify the noun. There is no agreement on adjectives and they can be repeated. Superlative adjectives display a behavior that is distinct from other adjectives in Persian. In contrast to the other adjectivals, the superlatives occupy the position preceding the noun.

**Adverb.** Adverbs may appear in the Adjectival Phrase (AP) preceding the adjective; they cannot be used without the latter in a NP.

**Pronoun.** Pronouns include personal as well as quantifying pronouns (**QPronouns**) such as *everyone* and *someone*. Pronouns usually appear in the position of the *possessor*, and they can form a whole NP on their own. Note that personal pronouns can appear either as separate lexical elements or as morphemes on the noun or adjective.

**Proper Noun.** Proper Nouns usually form an NP on their own. They usually occupy the position of the *possessor* noun.

**Determiner.** The Determiner precedes the head noun. The determiners are ‘*in*’ (*this*), ‘*ân*’ (*that*), ‘*har*’ (*each/any*).

**Numeral.** These consist of cardinal or ordinal numbers. The indefinite article ‘*yek*’ (*a/an, one*), which appears before the head noun, is treated as a Numeral. There are three different ordinal types in Persian and they all behave differently from the cardinal numbers, displaying properties similar to the superlative adjective.

**Unit.** The units are used in numeral constructions and usually follow the Numeral itself. These are lexical words such as ‘*hezâr*’ (*thousand*) or ‘*milyon*’ (*million*).

**Classifier.** Classifiers usually follow the numeral and classify the head noun by indicating its kind or type.

A numeral can appear without a classifier but a classifier can never appear without a numeral.

**Quantifier.** Quantifiers with meanings of *some* (e.g., ‘*ba’zi*’), *every* (e.g., ‘*hame*’), *any*, *no* (e.g., ‘*hich*’).

**Title.** Titles are forms of address preceding proper names.

**Infinitival.** The infinitival is used as a nominal within an NP. The infinitival, by virtue of being a verbal category, can appear with complements. The whole verbal structure is then nominalized and used in the NP.

**Conjunction.** Conjunctions can relate parts of the AP or NP.

**Article.** There is no definite article in Persian, only an indefinite. This article appears only as an affix attached to the noun or adjective.

## 2.2 Simple Noun Phrase

Simple NPs can also be divided into three distinct groups. The structure of the noun phrase with a nominal head (referred to here as “standard NP”) differs from that of the infinitival constructions, hence we will describe each NP type separately. Furthermore, the elements that can act as the *possessors*, such as proper names and pronouns, are listed as separate noun phrases.

### 2.2.1 Standard NP

The head Noun is followed by the *modifiers*, which usually consist of an Adjectival Phrase (AP) construction. There could be several *modifiers* in a Noun Phrase. The elements preceding the head noun are the determiner, the numeral constructions and the quantifiers. Although adjectives always follow the noun, the superlative adjective can only appear before the head. Numerals, quantifiers and superlative adjectives are in complementary distribution; if one of these elements is present, the others cannot appear within the NP. Since complementary distribution usually indicates that the lexical elements occupy the same position, the numeral, quantifier and superlative constructions are all placed under the *specifier* category.

The relative ordering of the constituents of the simple NP is as follows:

**NP =** *determiner specifier head modifier*

where the *head* is a Noun and the parts of speech or phrases that can appear in each of the other categories are as shown below. Note that all the constituents, with the exception of the head noun, are optional.

***determiner:*** Determiner

***specifier:*** Numeral (Unit) (Classifier)<sup>1</sup>  
Numeral [Ordinal]  
Adjective [Superlative]  
Quantifier

***modifier:*** (Adverb) Adjective  
***Note:*** Modifiers may be recursive

---

1. Brackets indicate optionality.

The example in (1) represents a simple Noun Phrase where CL stands for Classifier and EZ for the ezafe morpheme.

- (1) *in do tâ ketâb-e kohne*  
 this two CL book-EZ old  
 ‘These two old books’

### 2.2.2 Infinitival Noun Phrase

The head of the noun phrase can be an infinitival verb. These NPs are very similar to the gerundive constructions in English. Note that the infinitive head can be either a simple verb or a light verb, and it can appear in a predicate construction or with an adverbial.

The NP in (2) is an example of a predicative construction formed with the verb *budan* “to be”. The predicate element *zan* “woman” is placed before the verb. The objects of the verb, however, become arguments of a genitive (or possessive) construction as shown in the example in (3). The infinitival verb *koshtan* “to kill” carries an ezafe morpheme linking it to the object of the verb *shir* “lion”. Note that a similar relation is present in the English translation for this example, in which the object of the gerundive (i.e., “lion”) is linked to the deverbal noun by an “of” construction.

- (2) *zan budan-ash*  
 woman be-her  
 ‘her being a woman’

- (3) *koshtan-e shir*  
 kill-EZ lion  
 ‘the killing of a lion’

In our analysis, the NP structure in (2) is analyzed in the simple NP rules, as an infinitival NP. The structure in (3), however, is treated as a complex NP, formed from the concatenation of two distinct simple NPs. Complex NPs are discussed in Section 2.3.

The structure of the Infinitival Noun Phrase constituents can be described as either  
**NP = *predicate head*** or **NP = *adjunct head***

### 2.2.3 Possessor Noun Phrase

Possessive pronouns and proper names follow the head noun in Persian, as exemplified below:

- (4) *ketâb-e man*  
 book-EZ 1sg-pronoun<sup>2</sup>  
 ‘My book’

- (5) *ketâb-e dâryush*  
 book-EZ Dariush  
 ‘Dariush’s book’

2. Since there is no Case in Persian, the surface form of the pronoun is always the same whether it is used in a subject, object or possessive context.



Pronouns and proper names mark the boundary of the Noun Phrase (i.e., no other NP element can appear to the right of pronouns and proper names), hence they are often included as the last element in the Persian NP. Since these lexical elements can also be heads of their own NP (i.e., NPs can consist of simply a pronoun or a proper name), we will treat them as separate noun phrase constructions<sup>3</sup>. Section 2.3 will explain how these phrases are used to form bigger NPs. The possessor NP can consist of any of the following lexical elements:

- NP = *personal pronoun*  
 NP = *quantifier pronoun*  
 NP = (*title*) *proper name*

## 2.3 Complex Noun Phrase

The complex noun phrase is the equivalent of the genitive or possessive constructions in English, such as “Mao’s red book”, “her mother’s hat” or “the syntax of noun phrases”. In English, the link between the two nouns is marked by “ ‘s” (e.g., Mao’s) or the preposition “of”. In the case of pronouns, the latter appear in their genitive form (e.g., her). Other languages use Case to indicate the link between noun phrases. For instance, in the following Armenian example, each noun is joined to the following one by the Genitive morpheme “-i”<sup>4</sup>.

- (6) *yerk’r-i mayrakaghak-i poghots-ner-e*  
 country-Gen capital-Gen street-Plur-Nom  
 ‘the streets of the country’s capital’

The element joining the Persian noun phrase constituents to each other is the *ezafe* suffix. The *ezafe*, however, is usually pronounced as the short vowel /e/ and is therefore not marked in written text. The result, in Persian written text, is a series of consecutive nouns without any overt links or boundaries as shown in the example in (7) transcribed as it appears in Persian text (i.e., without short vowels). The actual pronunciation for this example is given in (8); the *ezafe* morpheme is represented by the *-e* following the first three nouns, linking each one to the following constituent. Note that the last constituent in the NP does not carry the *ezafe* suffix, thus marking the end boundary of the noun phrase.

- (7) *ktab dvst pdr daryvsh*  
 book friend father Dariush  
 ‘Dariush’s father’s friend’s book’

- (8) *ketâb-e dust-e pedar-e dâryush*

In this example, each noun forms a simple NP which then join together to form the complex NP given in (7). The lack of Case and agreement, as well as overt linking morphemes, coupled with a verb-final word order, can make the computational parsing of Persian NPs extremely ambiguous. In the next section, we will present possible boundary markers or joining elements in Persian that can help resolve some of the parsing ambiguities.

3. There are no capital letters in Persian, hence proper names are not easily differentiated from nouns.

4. *Gen* stands for Genitive Case, *Nom* for Nominative Case, *Plur* for the Plural morpheme.

### 3

---

## Noun Phrase Boundaries

### 3.1 Lexical Categories

The constituent ordering for the simple standard NP, discussed in Section 2.2, already points to some of the beginning boundaries of the Persian NP. Hence, the determiner, if present, is the first element of the noun phrase. If there is no determiner, the specifier is the first element. It was also mentioned, in Section 2.3, that the possessor elements constitute the boundaries of the complex noun phrase. In other words, if a simple nominal NP is followed by a possessor NP structure, the two NPs can join to form a bigger NP, but no other NP element can join to the right of this newly formed complex NP<sup>5</sup>.

Consider the sentence in (9) with its corresponding noun phrase boundaries as shown on the gloss in (10).

- (9) *bh gft̄h ayn xbrḡzary vzyr xarJh âyndh kshvr banv âlbrayt ast*  
according to this news agency minister foreign future country lady Albright is  
'According to this news agency the future foreign minister of the country is Lady Albright.'

- (10) [according to this news agency]NP [minister foreign future country]NP [ladyAlbright]NP is

In this example, there are eight NP constituents between the preposition “according to” and the final verb. The NP boundary can, in principle, fall after any of the nouns in this sentence, which leads to a very high parsing ambiguity. Now compare (9) to the sentence below with proper nouns:

- (11) *bh gft̄h xbrḡzary fransh vzyr xarJh âyndh âmryka banv âlbrayt ast*  
according to news agency France minister foreign future United States lady Albright is  
'According to France's news agency the future foreign minister of United States is Lady Albright.'

- (12) [according to news agency France]NP [minister foreign future U.S.]NP [lady Albright]NP is

In this case, the proper names can be used to detect the final boundaries of the noun phrases, thus analyses joining *Amryka* (US) and *banv* (Lady), or linking *fransh* (France) and *vzyr* (minister) will not be formed.

---

5. There are cases as in (i), where the proper name can be modified and even joined to the right by another possessor (here, a pronoun) but such instances seldom occur in written text.

(i) *shirâz-e zibâ-ye mâ*  
Shiraz-ez beautiful-ez our  
'our beautiful Shiraz'

### 3.2 Indefinite Article / Enclitic Particle (IE)

Although the indefinite article and the enclitic particle (which links the head noun to the following relative clause) have different syntactic functions, they have the same surface representation and cannot be differentiated in morphology. This morpheme can appear on a noun (as in examples (13) and (14)) or on an adjective (as in (15)). If there is more than one adjective, the IE will appear on the head adjective (the last one) as shown in (16).

- (13) *nkth~ay* [nokte-i]  
point-IE  
'a point / an issue'
- (14) *nkth ~ay mhm* [nokte-i mohem]  
point-IE important  
'an important point'
- (15) *nkth mhmy* [nokte-ye mohem-i]  
point important-IE  
'an important point'
- (16) *nkth Jalb v mhmy* [nokte-ye jâleb va mohem-i]  
point interesting and important-IE  
'an interesting and important point'

In all of these instances, the presence of the IE marks the NP boundary, in the sense that no other NP element can follow the noun in (13) or the noun-AP combination in (14) through (16). This is exemplified below:

- (17) *mrđmy Azadh rvabT aqtSqdy ayran ra mhm mydannđ kh ...*  
people-IE noble relations economic Iran OBJ important know-3pl that ...  
'(Those) noble people consider Iran's economic relations important who...'
- (18) [people-IE noble]NP [relations economic Iran OBJ]NP important know that ...

Since the noun *mrđm* appears with the enclitic affix *y*, the simple NP consisting of *mrđmy Azadh* is not allowed to join to the NP to the right.

### 3.3 Personal Pronoun Clitics

Instead of appearing as a lexical pronoun, the possessive pronoun may be cliticized onto the rightmost constituent of the simple NP as shown in the two examples below. When the clitic is present, it marks the end boundary and the simple NP can not join to the following nominal element to form a complex NP.

- (19) *ktabsh* [ketâb-ash]  
book-Clitic  
'his/her book'
- (20) *ktab khnh~ash* [ketâb-e kohne-ash]  
book old-Clitic  
'his/her old book'

The sentence below shows how the clitic is used to denote the end boundary of the Noun Phrase, thus not allowing it to join to the following element.

(21) *hmsayganman ktab khnh shma ra brdashtnd*  
 neighbor-Plur-Clitic(1pl) book old you(2pl) OBJ took-3pl  
 ‘Our neighbors took your old book.’

(22) [neighbor-Plur-Clitic]NP [book old your OBJ]NP took.

### 3.4 Ezafe Morpheme

The ezafe morpheme does not mark the end boundary of a Noun Phrase but rather the lack thereof, since the ezafe is used to join the head of a NP to the constituents following it. As already mentioned, the ezafe is rarely written in Persian text since it is a short vowel. When it appears after a vowel, however, it has the surface form *y*. In these cases, the ezafe can be used to indicate that the simple NP should be joined to the following nominal phrase.

In the sentence given in (23), the adjective *zyba* (beautiful) appears with an overt ezafe morpheme, which indicates that the simple NP *zn zybay* (beautiful woman/wife + EZ) should be joined to the following NP (Dariush) thus forming the complex noun phrase *zn zybay daryush*. In other words, the NP boundary can *not* be at this location.

(23) *zn zybay daryvsh vard shod [zan-e zibâ-ye dâryush vâAred shod]*  
 wife beautiful-EZ Dariush entered  
 ‘Dariush’s beautiful wife entered.’

(24) [wife beautiful Dariush]NP entered.

In the second example below, on the other hand, the adjective *zyba* (beautiful) does not carry the ezafe suffix. Note that since this word ends in a vowel, if the ezafe were present, it would have appeared in its overt form *y*, hence we can conclude with certainty that the ezafe is not available. The absence of the ezafe indicates that a boundary should be set following the adjective thus forming two separate noun phrases as shown in (26).

(25) *zn zyba daryvsh ra shnaxt [zan-e zibâ dâryush râ shenâkht]*  
 woman beautiful Dariush OBJ recognized  
 ‘The beautiful woman recognized Dariush.’

(26) [woman beautiful]NP [Dariush OBJ]NP recognized.

Unfortunately, in most cases, the ezafe is an unwritten short vowel and cannot provide any information as to the boundaries of the NP. In our grammar, we treat such cases as having an “Undefined” ezafe affix. When the ezafe is “Undefined”, it is treated as both a boundary and as a single NP resulting in ambiguous parsing.

### 3.5 Combination of Boundary-Marking Features

The following table presents the possible cooccurrence for the ezafe, indefinite/enclitic and the pronominal clitic morphemes. The combination of these features is used in certain rules in the

syntactic parser. So, for instance, if the *ezafe* is marked as *Undefined*, the other features can only be *False*.

	<b>combination 1</b>	<b>combination 2</b>	<b>combination 3</b>	<b>combination 4</b>	<b>combination 5</b>
<i>ezafe</i>	True	False	False	False	Undefined
<i>clitic</i>	False	True	False	False	False
<i>indefinite/ enclitic</i>	False	False	True	False	False

## 4

# Bolero Grammar

---

This section introduces a few sample NP rules from the Bolero syntactic grammar. These rules demonstrate how the information from the structure and the boundary-marking elements of the NP are incorporated within the grammar.

The presence of a boundary marker, such as the indefinite/enclitic morpheme, is denoted by a feature on the NP feature structure rules called *boundary*. When a boundary marker (e.g., clitic or IE) is encountered, the value for this feature is set to True. The True value indicates that the NP has reached a boundary and cannot join to the following constituent to form a bigger noun phrase. If a boundary-marking morpheme was not found on the NP constituents, the value is set to False. In such cases, the NP is free to join to the next element. In certain cases, as when the presence or absence of an *ezafe* morpheme can not be determined, the boundary is set to “Undefined”, in which case the NP may or may not join to the constituent following it.

Consider the rule `NounBarIndefinite` given below. This rule contains a left-hand side (*lhs*) and a right-hand side (*rhs*) as in rewrite rules.

```
NounBarIndefinite = per.Rule[
  lhs: per.NounBar[
    head: #head,
    boundary: True],
  rhs: <:
    #head= per.Noun[infl.indefiniteEnclitic: True]]
  :>
];
```

The right-hand side of this rule is satisfied if an entry with type `Noun` is recognized, which also carries an indefinite/enclitic morpheme. As can be seen in the left-hand side of this rule, this nominal element is tagged as the head of the N' and the value of the *boundary* feature is set to True. The boundary value is transferred up when the higher NP level is formed as shown below in the `NPo` rule.

The `NPo` is the feature structure forming a standard simple NP. It contains all of the constituents that could constitute the standard noun phrase as described in Section 2.2.1 “Standard NP”. Each constituent on the *rhs* points to the elements in the feature structure in the *lhs* of the rule, through a variable (marked by the pound sign #). As mentioned, the boundary value that was set for the `NounBar` (N') is also transferred up to the `NPo` structure.<sup>6</sup>

```
// NPo --> Det? Spec? N' where N' --> Noun Adj?
NPo = Rule[
  lhs: per.NounPhraseZero[
    determiner: #det,
    specifier: #spec,
    head: #head,
    modifier: #mod,
    boundary: #bnd],
```

6. The value of the variable `#bnd` in this rule is set to `Top`, indicating that the actual value can be anything, i.e., in the case of the boundary feature, the value could be set to True, False or Undefined. This value is then percolated up to the *lhs* of the rule.

```

rhs: <:
  "optional" #det= per.Determiner
  "optional" per.Specifier[specType: #spec = Top]
per.NounBar[
  head: #head = Top,
  modifier: #mod = Top,
  boundary: #bnd = Top]
:>
];

```

The complex noun phrase, which consists of two or more simple NPs, is formed using the recursive rule called `complexNP`. The right-hand side of this rule looks for an `NP0` structure followed by a Noun Phrase feature structure<sup>7</sup>. This construction could be exemplified with the noun phrase sample in (23) *zn zybay daryvsh* (woman beautiful-Ez Dariush), in which the simple NP (or `NP0`) “woman beautiful-Ez” and the proper name NP “Dariush” join to form a bigger NP. What should be noted is that the right-hand side of this rule is satisfied only if the boundary value is set to `False` or to `Undefined`. Hence, if the boundary value is `True`, such as when an IE morpheme is encountered, the complex noun phrase will not be formed.

```

complexNP = per.Rule[
  lhs: per.NounPhrase[
    head: #np1,
    possessor: #np2],
  rhs: <:
    #np1= per.NounPhraseZero[
      boundary: FalseOrUndefined]
    #np2= per.NounPhrase
  :>
];

```

---

7. The Noun Phrase or NP structure is used to indicate noun phrases that have already reached a boundary on their rightmost element.

## 5

### Relative Clause

---

Relative Clauses are used to give further information about a nominal element, such as in the English sentence “The man, whom I met yesterday, has had an accident.”, where “whom I met yesterday” represents a relative clause providing further information about “the man”. In Persian, relative clauses are usually introduced by the relativizer *kh* [*ke*] (that), which is used regardless of the animacy, gender or function of the head noun. In nonrestrictive relative clauses, the head noun often carries the Enclitic morpheme which links it to the following relative clause. In these instances, the head noun is usually interpreted as a definite<sup>8</sup>.

The relative clause construction is similar to English: The head noun is followed by the relativizer (*kh* in Persian), which is then followed by the clause that relates to the head noun, as shown below:

*head noun* [ **'kh'** [ *Clause* ] ] ...

In certain cases, the relative clause can be separated from the head noun by the verb of the sentence. In addition, several relative clauses could follow a head noun. As mentioned above, the relativizer *kh* does not vary depending on animacy or function of the head noun; in other words, relative pronouns such as “who”, “which”, “whom” do not exist in Persian. It is also not possible to precede the relativizer by a preposition as in the English examples “to whom”, “in which”.

If the head noun is the subject or direct object of the relative clause, it is often left as a gap as shown in the examples in (27a) and (27b), respectively. Note that the subject in the clause (*tv* “you”) is optional, since Persian is a pro-drop (i.e., optional subject) language.

- (27) a. *zn-y kh ktab myxvand [zani ke ketâb mikhânad]*  
woman-Encl that book read(pres/3sg)  
‘The woman who reads books’
- b. *zn-y kh tv myshnasy [zani ke to mishenâsi]*  
woman-Encl that you know(pres/2sg)  
‘The woman that you know’

In certain instances, however, even if the head noun is the subject or direct object of the relative clause, it may be replaced by a pronoun in the clause it originated from. In (28), the head noun *plak kvchk* (small plaque) is the subject of the relative clause; it is substituted by the resumptive pronoun *ân* (it). The use of the resumptive pronoun usually occurs when the head noun is separated from the relative clause by an intervening verb. In this example, the verb *py brdh and*

---

8. It is not always a definite, however, as shown in the example in (28), where the head noun is followed by an enclitic, but it is translated as an indefinite in English. It may be more of a specificity marker than a definiteness marker. This point requires further research.



(have found) precedes the relative clause.

- (28) *danshmndan bh plak kvchk-y dr mQz py brdhand kh An nyz takvvn nashnaxth mandh bvd.*  
 scientists to plaque small-Encl in brain found that it also until now unknown had remained  
 ‘Scientists have found a small plaque in the brain that until now had remained undiscovered.’

When the head noun is the indirect object or is extracted from a Prepositional Phrase adjunct in the clause, a resumptive pronoun is used. In other words, the position from which the head noun originates is substituted by a pronoun that agrees with the head noun. This is exemplified in the three NP cases below:

- (29) *ayn bchh-ha kh az Anha Adrs myprsydy [in bachehâ ke az ânâ âdres miporsidi]*  
 this kid-Plur that from them address ask(imp/2sg)  
 ‘These kids from whom you asked for the address’

- (30) *shhr-y ke dr An tZahrat shdh bvd [shahri ke dar ân tazâhorat shode bud]*  
 city-Encl that in it demonstrations become(pluperf/3sg)  
 ‘The city in which demonstrations took place’

- (31) *zn-y kh bray-sh ktab xrydy [zani ke barâyash ketâb kharidi]*  
 woman-Encl that for-Clitic(3sg) book buy(past/2sg)  
 ‘The woman for whom you bought a book’ or  
 ‘The woman that you bought a book for’

In (29), the head noun *ayn bchh-ha* (these kids) is the indirect object of the clause; it is extracted from the PP complement of the verb “ask”. As the example shows, the preposition *az* (from) is left behind in the relative clause, and the head noun is replaced by a pronoun *Anha* (they/them). A similar example is given in (30) with an inanimate head noun. In (31), the head noun *zn* (woman) is also extracted from the PP complement of the clausal verb. In this instance, however, the head noun is replaced by a clitic pronoun *-sh* (him/her), which appears attached on the preposition *bray* (for). The word for word gloss for this example is then *woman that for-her book (you) bought*.

If the head noun of the relative clause is the object of the main sentence, then it may appear with the object marker *ra*, as shown in the following sentence. Note that the head noun *flsTynyany* receives an object marker, even if it is the subject of the relative clause.

- (32) *flsTynyany-y ra kh dr xyabanha tZahrat mykrndnd kshnd*  
 Palestinians-Encl OBJ that in streets demonstration do(imp/3pl) kill(past/3pl)  
 ‘(They) killed the Palestinians who were demonstrating in the streets.’

The examples below show a head noun separated from the relative clause by an intervening verb (also see example (28)) and an intervening adverb, respectively.

- (33) *nxstyn kar-y ast kh dr ayn kshvr bh nfe fransvyha anJam shdh ast*  
 first work-Encl is that in this country to benefit french-Plur perform(passive/perf/3sg)  
 ‘It is the first thing that has been performed to the benefit of the French in this country.’

- (34) *bradr av nyz kh bshdt eSbany shdh bvd nqshh qtl Jan ra kshyd*  
 brother his also that intensely angry had become plan murder John OBJ pulled  
 ‘His brother also who had become intensely angry planned John’s murder.’

## 6

---

### **Conclusion**

This report describes the structure of the Noun Phrase in Persian and explains how certain morphological and syntactic features could be helpful in determining boundaries of Noun Phrases in Persian. These constraints, when incorporated within the syntactic grammar, can reduce the number of parses produced during analysis. The way in which these boundary markers were incorporated within the Bolero grammar, used in specifying the syntactic structure of Persian, is also discussed. The final section covers the relative clause constructions in Persian.

## Appendix

Table 1: Romanization used in the report

Persian Letters	Romanization	Pronunciation
alef with madd	A	f <u>ather</u>
alef	a	<u>and</u> or <u>bed</u> or <u>so</u>
be	b	<u>boy</u>
pe	p	<u>pool</u>
te	t	<u>toy</u>
se	s	<u>sun</u>
jim	J	<u>Joe</u>
che	ch	<u>church</u>
he	H	<u>horse</u>
khe	x	<i>similar to German buch</i>
dal	d	<u>dog</u>
zal	z	<u>Zorro</u>
re	r	<i>similar to Spanish "r"</i>
ze	z	<u>Zorro</u>
zhe	j	<u>mirage</u>
sin	s	<u>sun</u>
shin	sh	<u>shoe</u>
sat	S	<u>sun</u>
zat	Z	<u>Zorro</u>
ta	T	<u>toy</u>
za	Z	<u>Zorro</u>
eyn	e	<u>and</u> or <u>bed</u> or <u>so</u> or <u>uh</u> oh (glottal stop)
gheyn	Q	<i>similar to French "r"</i>
fe	f	<u>fun</u>
ghaf	q	<i>similar to French "r"</i>
kaf	k	<u>kite</u>
gaf	g	<u>great</u>
lam	l	<u>love</u>
mim	m	<u>Mary</u>
nun	n	<u>nun</u>
vav	v	<u>very</u> or <u>food</u>
he	h	<u>horse</u>
ye with hamze	i	<u>you</u> or <u>uh</u> oh (glottal stop)
ye	y	<u>you</u> or <u>sea</u>
short space (marking a final form character)	~	