

Noun Phrase or Compound Noun? An Investigation of N + A and N + N Boundary Cases in Persian

Shahram Peyman

University of Isfahan, Isfahan, Iran

Vali Rezai

University of Isfahan, Isfahan, Iran

Abstract—The present study seeks to investigate the demarcation between noun phrases (NPs) and compound nouns (CNs) in Persian at the syntax-morphology interface. This objective is accomplished through the examination of two most complex nominal patterns, viz. N + A, N + N, with special focus on boundary cases, i.e. the intermediate constructs which possess some properties of both NPs and CNs simultaneously and thus demonstrate contradictory reactions to the various NP-CN demarcation criteria. The results indicate that boundary cases ensue from partial syntactic erosion of NPs through pure lexicalization, whereby NPs turn into CNs without center-switching or category change. This study also shows that almost all boundary cases have no potential for syntactic modification of their elements. It is further demonstrated that N + A and N + N boundary cases are endocentric, head-initial constructs with optional or obligatory internal inflection as well as *Ezafe*. Syntactic modifiability is also introduced as the most efficient NP-CN demarcation criterion in Persian since it is the first property lost in NP lexicalization process.

Index Terms—compound noun, noun phrase, demarcation criteria, lexicalization, syntactic erosion

I. INTRODUCTION

One of the differences between traditional grammars and modern linguistic theories resides in a modular approach to the study of language. This approach advocates the separation of autonomous language modules as syntax, semantics, phonology and morphology (see Chomsky, 1957, 1965, 1975, 1980, 1981, 1986; Fodor, 1983, 1989; Newmeyer, 1983). Chomsky's generative grammar, for instance, has always been faithful to this modular approach (Dabir-Moghaddam, 1999).

The autonomy of language modules, however, does not mean that the interactions and interfaces among these modules should be neglected. Each one of these modules enjoys an independent systematic structure, while retaining interfaces with other modules at the same time. Ultimately, the modules of language have mutual impacts upon one another.

When it comes to the interface between morphology and syntax, as Booij (2005) argues, one important issue that must be addressed is word-phrase demarcation. The main concern here is to determine when a multi-morphemic sequence is a word and when it functions as a phrase.

The present study deals with syntax-morphology interface in Persian as far as word-phrase demarcation in nominal constructs is concerned. An interesting phenomenon in Persian is the existence of nominal constructs which seem to have a hybrid nature, reflecting characteristics of both CNs and NPs simultaneously. This study aims to examine these hybrid constructs, referred to as boundary cases, and seeks to assess and analyze various demarcation criteria in order to provide the most effective ones to distinguish NPs from CNs in Persian.

II. WORD-PHRASE DEMARCATION CRITERIA

Among the studies conducted on word-phrase demarcation, one can find a variety of criteria, the most important of which will be reviewed in this section.

Katamba (1993) considers compound word as a word which consists of at least two bases that can each stand as an independent word elsewhere. He explains the fact that orthographic conventions are not much helpful in distinguishing between compound words and phrases. He argues further that English spelling does not have a uniform set of conventions, so that even linguists such as Rohrer (1974), Aronoff (1976) and Bauer (1983) may differ in their spellings of the term "word formation".

Katamba also addresses phonological criteria and explicates 'accent subordination' in Bloomfield's view (Bloomfield, 1933). He maintains that accent subordination is a useful criterion; however, it does not apply to all compound words, as Marchand (1969) asserts.

He then elaborates on the concept of 'listeme', as explained by Di Sciullo and Williams (1987), and argues that words have to be listed in the lexicon since they have idiosyncratic properties which are not subject to general principles, so native speakers have to memorize them. In contrast, syntactic phrases, which are formed and analyzed on the basis of given rules, need not be listed in the lexicon. Nevertheless, Katamba rejects 'listedness' as the one and only criterion to differentiate between compound words and phrases. Katamba then mentions the semantic criterion of 'compositionality' and states that the less compositional the nature of a linguistic form is, the more likely it is for this form to be listed.

In Katamba's view, the main distinction between compound words and phrases lies in the fact that the internal structure of compound words, whatever form it may have, is inaccessible to syntactic rules and these rules, which operate at the level of phrase, do not act on words. This approach, known as 'lexicalist hypothesis', has been adopted by Chomsky (1970), Anderson (1988) and Williams (1981) as well.

Booij (2005) asserts that the distinction between compound words and phrases is not always easy to establish owing to two reasons: the similar labeling function that both of them can perform and their close similarity as a result of the historical derivation of compound patterns from phrasal word combinations.

Booij holds that the most significant wordhood criterion is 'lexical integrity'. Following Anderson (1992), Booij defines this principle as below:

'The syntax neither manipulates nor has access to the internal form of words'

In fact, this principle appears to make a claim similar to that of 'lexicalist hypothesis' mentioned above. Now a prominent question arises as to the implications and consequences of lexical integrity. One of these implications, according to Booij, is that inflectional rules, which are governed by syntax, cannot apply to the internal elements of words. In fact, words cannot have internal inflection. Another consequence of this principle is that the elements within a compound word cannot be independently modified by means of modifiers. A third corollary of lexical integrity is the inseparability of the compound elements. Therefore, a multi-morphemic sequence is a phrase if its constituents can be separated.

Apart from lexical integrity, Booij points out the phonological criterion of stress and explains that compound words of A + N structure take the main stress on their first constituent and A + N phrases have it on their second. Booij also enumerates semantic criteria such as semantic idiosyncrasy, yet he does not consider them to be as effective and helpful as formal criteria.

In his analysis of separable complex verbs in Dutch, Blom (2005) refers to lexical integrity as discussed by Lapointe (1980) and Bresnan and Mchombo (1995) and emphasizes constituent inseparability and inaccessibility to syntactic rules on the part of compound words. He recognizes constituent separability of Dutch complex verbs as the major reason against regarding them as compound words. It seems that lexical integrity is the most crucial criterion for word-phrase demarcation in Blom's view.

Haspelmath (2002) also mentions the criterion of lexical integrity, but discusses the challenges and counterexamples as well. Under the topic of noun incorporation, he maintains that in some languages like Southern Tiwa verbs agree with the incorporated nouns. Agreement is a syntactic rule, so this case exemplifies the intrusion of syntactic rules into word boundary. He also points to some other languages like Greenlandic Eskimo, in which the incorporated noun has unincorporated modifiers.

In a number of studies in Persian, the issue of word-phrase demarcation has been addressed directly or indirectly. Bateni (1994), who employs the Scale and Category framework in the description of Persian, first provides a structural definition of NP. It is in the explanation of NP constituents that he is confronted by the problem of NP versus CN. Bateni states that it is sometimes difficult to distinguish between prefix + base and pre-modifier + head constructs because they lie somewhere between the two poles. He then proposes a continuum with the two constructs at the extreme points and intermediate instances in between.

Afrashi (2010) examines morphosyntactic constructs in Persian at the syntax-morphology interface. She regards these constructs as morphological units of compound or derived-compound type which have a syntactic relation between their constituents. Each such construct is named after its syntactic head. Subjective, objective, Ezafe, attributive, appositive, complement, adverbial, conjunctive and numerical compounds are described as different forms of morphosyntactic constructs.

Meshkatod Dini (2002) defines compound word as a word composed of two lexical morphemes. He then attempts to present various compounding patterns in Persian. He also explicates the structure of different phrases. However, he does not elaborate on the demarcation of compound words and phrases.

Shaghghi (2012) lists four differences and five similarities between compound words and phrases. The first difference relates to the constructional restrictions on compounding. It is not possible to form compounds with all potential bases. Such limitations do not exist in the case of phrases. The second difference has to do with lexical integrity. Shaghghi explains that no syntactic or inflectional elements can be added between bases in compound words. The third one pertains to the feature of compositionality in case of phrases, while compounds do not enjoy such a characteristic. The last distinction is connected with the concept of head. She claims that compounds may lack head and if there is one, it takes all the derivational and inflectional affixes. In a phrase, however, there is always a head which takes inflectional affixes.

As to the similarities between compounds and phrases, Shaghghi asserts that both have constituent structure, which means a new constituent can be added to the old ones. The second similarity concerns the relation between elements in a compound and a phrase. Head-modifier and verb-argument relations as well as coordinate structure can be found in both compounds and phrases. As the third one, Shaghghi maintains that compounds and phrases both consist of two or more simple or complex words. The fourth likeness relates to the fact that some compounds may enjoy the feature of compositionality just like phrases. The final similarity in Shaghghi's view is that endocentric compounds have heads like phrases.

III. NP-CN DEMARCATION IN PERSIAN

In Persian, there exists a multiplicity of patterns forming a large number of nominal constructs which are unambiguous in nature as they exhibit consistent results when NP-CN demarcation criteria are applied to them and thus are undoubtedly either CNs or NPs. However, there are boundary cases which react inconsistently to demarcation criteria and entail either ambiguous or even contradictory results.

In this section of the present study, we shall address ourselves to the demarcation of N + A and N + N NPs and CNs in Persian, the distinction of which is not always easy to establish, and the evaluation of the relevant criteria described in the literature. In doing so, particular attention will be paid to boundary cases and the theoretical implications thereof. In the forthcoming discussion, the morphological criterion of lexical integrity, the phonological criterion of stress and the semantic criterion of listedness as well as semantic idiosyncrasy will be considered. The authors will also propose other criteria, some of which are exclusively applicable to Persian.

N + A and N + N are the most complex nominal patterns as they can form constructs of profoundly different nature. They share the characteristic of forming both CNs and NPs. It seems that the CNs formed by these two patterns are derived from their respective NPs through the process of lexicalization, whereby NPs undergo **syntactic erosion** and lose their syntactic properties to become CNs. What contributes to the formation of boundary cases is the fact that this syntactic erosion is not always complete. **Partial syntactic erosion**, which refers to the loss of only some syntactic properties and the retention of some others, gives rise to the formation of hybrid constructs that partly enjoy the features of both NPs and CNs, hence demonstrating contradictory reactions to demarcation criteria. Through the analysis of these two patterns and the comparison of the constructs which they create, it is possible to have deeper insights into the nature of noun compounding process and the emergence of boundary cases in Persian. The in-depth investigation of the aforementioned patterns will follow.

A. The N + A Pattern

This pattern, which is very common in Persian, can create both NPs and CNs. The NPs have the structure of head (N) + post-modifier (A). The CNs, however, might be of two types: exocentric and endocentric. Each one of these structures will be reviewed in the following discussion.

1. N + A as NP

In order for N + A to be an NP, at least three conditions must hold true:

a. The existence of Ezafe between the two elements: in Persian, Ezafe (/ -e/) necessarily occurs between head and nominal or adjectival post-modifiers. If N + A is to be an NP with the structure of head + post-modifier, Ezafe inevitably appears between them.

b. Internal inflection: in NPs, the head N is inflected independently of its modifiers. For instance, the plural affix (i.e. /hâ/ or /ân/) as an inflectional morpheme is directly added to the head N. In N + A NPs, only N can take the plural affix.

c. Syntactic modifiability: in NPs, both the head and its modifiers can take their own modifiers independently. For example, A as one modifier of N can have an ADV as its modifier. In N + A NPs, both elements can be expanded through their own modifiers.

In order to have a more profound understanding of the conditions above, we will analyze an instance of N + A NP in Persian. The example (1a) below is an NP consisting of N + A.

- (1) a. gol-e ziba
flower-EZF beautiful
"a beautiful flower"
- b. gol-hâ-ye ziba
flower-PL-EZF beautiful
"beautiful flowers"
- c. *gol-e ziba-hâ
- d. gol-e xošbu-ye besiyâr ziba
flower-EZF fragrant-EZF very beautiful
"a very beautiful, fragrant flower"

There is Ezafe between the head N and A, as can be seen in the phrase. (1b) is the only acceptable plural form of this phrase, where the plural affix /hâ/ is added to the N (/y/ is merely an intervening consonant that prevents the adjacency of two vowels). Note that the addition of plural affix to A leads to the ungrammatical string (1c). Also, it is possible to modify both N and A in this phrase, as in (1d).

2. N + A as CN

Now we should consider the status of these three features (i.e. Ezafe, internal inflection, and syntactic modifiability) in the case of CNs.

a. Ezafe is a bound morpheme which occurs on the outermost layer of the word after all derivational and inflectional affixes and links a nominal or adjectival post-modifier to the head. In fact, it is found on word boundary. Therefore, CNs are not expected to have Ezafe inside them. This conclusion has some complexities, however. As will be seen in the remainder of the discussion, Ezafe shows a much more complicated behavior in the case of CNs than expected.

b. As CN is one word and must enjoy the feature of lexical integrity, the inflectional plural affix is not expected to enter its boundary. In fact, the N should not be independently pluralizable. This feature is referred to as ‘no internal inflection’.

c. On the basis of the reason mentioned above with regard to lexical integrity, it is anticipated that the elements in CN will not be syntactically modifiable. This characteristic is called ‘no syntactic modifiability’.

As stated before, the N + A pattern can form both exocentric and endocentric CNs. They demonstrate such distinct properties that they merit separate examination of their structure and formation. In the discussion below, each one of them will be analyzed separately.

A very common type of exocentric CNs formed by this pattern includes those cases which are generally known as ‘bahuvrihi compounds’. Booi (2005) considers bahuvrihi compounds as a subset of exocentric compounds. The construct in (2) below is an example of such forms:

- (2) a. *češm-e siyâh* (NP)
 eye-EZF black
 “a black eye”
 b. *češm siyâh* (CN)
 eye black
 “a black-eyed person”
 c. *češm siyâh* (CA)
 eye black
 “black-eyed”

The NP (2a) goes through the process of **center-switching** and is converted from an NP with the structure of head + post-modifier, which is inherently endocentric, into the N + A CN (2b), which is exocentric in this case. As a result of that, it loses the features of syntactic modifiability, internal inflection and Ezafe. The process of **category change** also affects this form and produces the compound adjective (2c) out of its respective CN. The authors would use the term **compound lexicalization** to refer to the process of lexicalization combined with center-switching or category change. The process described above is very productive in Persian and can be adopted in forming a large number of CNs. The CNs thus produced are not expected to bear any ambiguities in NP-CN demarcation due to two reasons: first, they lack Ezafe, which is a crucial element in NPs. Second, they are exocentric, whereas NPs are inherently endocentric. Owing to these two sharp distinctions, the N + A exocentric CNs and their respective NPs are easily distinguished.

The N + A pattern also creates endocentric CNs. The example (3a) below is one such CN:

- (3) a. *kâr dasti* (CN)
 work manual
 “(children’s) handicraft”
 b. *kâr dasti-hâ*
 work manual-PL
 “(children’s) handicrafts”
 c. **kâr-hâ-ye dasti*
 d. *kâr dasti-ye zibâ*
 work manual-EZF beautiful
 “(children’s) beautiful handicraft”
 e. **kâr-e zibâ-ye dasti*
 f. *kâr-e dasti* (NP)
 work-EZF manual
 “work done by hand”

Showing no capacity for internal inflection (3b vs. 3c) and syntactic modifiability (3d vs. 3e) and lacking Ezafe, this form appears to result from the lexicalization of the NP (3f) into a CN via the process of complete syntactic erosion.

3. N + A Boundary Cases

A large number of N + A endocentric constructs, however, display a rather unusual property, i.e. they have optional Ezafe, contrary to the expectations about CNs. This means that they can appear both with and without Ezafe. The example in (4a) represents this class of constructs:

- (4) a. *čerâq(-e) râhnamâ*
 light-EZF guiding
 “traffic light”

- b. čerâq-hâ-ye râhnamâ
light-PL-EZF guiding
“traffic lights”
- c. čerâq râhnamâ-hâ
light guiding-PL
“traffic lights”
- d. čerâq râhnamâ-ye bozorg
light guiding-EZF big
“a big traffic light”
- e. *čerâq-e bozorg-e râhnamâ

Along with this feature, they also enjoy another odd characteristic which seems to be closely connected to the status of Ezafe inside them. They have the feature of **double inflection**, which enables them to receive plural affix in two positions: between the N and the A as internal inflection, like NPs (in 4b), and at the end of the construct after the A, like CNs (in 4c). It should be noted that the N in such constructs does not have the capacity for independent syntactic modification (4d vs. 4e).

As shown above, such constructs bear contradictory features: while they exhibit optionality with regard to Ezafe and internal inflection, their elements are not syntactically modifiable. These are the boundary cases which look both like NPs and CNs simultaneously.

Still more intriguing are those endocentric constructs which appear to have obligatory Ezafe. (5a) below represents one such form:

- (5) a. sag-e ?âbi
dog-EZF aquatic
“beaver”
- b. sag-hâ-ye ?âbi
dog-PL-EZF aquatic
“beavers”
- c. *sag-e ?âbi-hâ
- d. sag-e ?âbi-ye kučak
dog-EZF aquatic-EZF small
“a small beaver”
- e. *sag-e kučak-e ?âbi

This construct must be a CN owing to semantic idiosyncrasy and listedness, but it is extremely interesting that this form, unlike other forms discussed so far, has obligatory internal inflection (5b vs. 5c). On the basis of this fact, it is expected to be an NP. On the other hand, the criterion of syntactic modifiability seems to verify its CN status (5d vs. 5e).

Such constructs form another group of intermediate structures between NPs and CNs. Whereas they enjoy Ezafe and internal inflection like NPs, they are also similar to CNs in that their elements are not apt for independent syntactic modification.

The last two types of construct (i.e. endocentric N + A forms with optional or obligatory Ezafe) seem to have derived from their corresponding NPs through the process of partial syntactic erosion, whereby only some syntactic properties are lost while the rest remain unchanged.

B. The N + N Pattern

The N + N pattern is by far the most complex pattern in terms of variety in the constructs which it forms. Like N + A, this pattern can create both NPs and CNs. The NPs are composed of head (N) + post-modifier (N). It must be noted that an N can never function as pre-modifier in an NP in Persian; therefore, the N + N NPs cannot be of the type pre-modifier (N) + head (N). The CNs formed on the basis of this pattern fall into two categories: exocentric and endocentric. What adds to the complexity of this pattern is that endocentric constructs are also divided up into two groups: head-initial and head-final. Below is a detailed discussion of the various constructs formed by this pattern.

1. N + N as NP

In order to count as an NP, a given N + N construct requires the same three conditions that were stipulated for N + A forms, namely obligatory Ezafe, internal inflection and syntactic modifiability of the constituent elements. As an instance of this pattern, the NP (6a) below can be considered:

- (6) a. divâr-e ?otâq
wall-EZF room
“the wall of the room”
- b. divâr-hâ-ye ?otâq
wall-PL-EZF room
“the walls of the room”
- c. divâr-e boland-e ?otâq-e bozorg
wall-EZF high-EZF room-EZF big
“the high wall of the big room”

This form consists of N + N. As can be seen, there exists the element of Ezafe between the two parts, which is obligatory. The plural form of this phrase is shown in (6b), in which the plural affix /hâ/ is added to the first noun, which is the head. Both nouns in this phrase can be expanded through modifiers on their own, as in (6c).

2. N + N as CN

Just like N + A CNs, N + N CNs are not expected to have Ezafe, internal inflection and independent syntactic modifiability, owing to the same reasons elaborated on in the previous section about the N + A pattern. It was stated earlier that both exocentric and endocentric CNs might be formed based on this pattern. Endocentric CNs in this pattern are also classified into two categories of head-initial and head-final constructs. These diverse forms will be examined one by one in the following parts.

The exocentric N + N CNs all share the features of “no syntactic modifiability” as well as “no internal inflection”, in accord with our expectations. However, Ezafe does not appear to act consistently in such constructs. There are exocentric CNs that may have optional or even obligatory Ezafe. To gain more profound insights into such forms, we will analyze one instance, i.e. the form in (7a).

- (7) a. *pomp-(e) benzin* (CN)
 pump-EZF gasoline
 “gas station”
 b. *pomp-(e) benzin-hâ/*
 pump-EZF gasoline-PL
 “gas stations”
 c. **pomp-hâ-ye benzin*
 d. *pomp-(e) benzin-e bozorg*
 pump-EZF gasoline-EZF big
 “a big gas station”
 e. **pomp-e bozorg-e benzin*
 f. *pomp-e benzin* (NP)
 pump-EZF gasoline
 “gas pump”

This form may optionally contain Ezafe. The phonological criterion of stress or juncture does not apply here as there is no difference between an NP and a CN in terms of these two features in this case. As far as semantic criterion of listedness is concerned, it seems that this form enjoys the characteristic of semantic idiosyncrasy and must thus be listed in the lexicon. However, Booij (2005) considers formal criteria to be superior to semantic criteria. Now we should see the consequences of the most important word-phrase demarcation criterion, i.e. lexical integrity. First, the form is tested for internal inflection. If the plural affix /hâ/ is added to the end of the construct after the second element, the well-formed construct (7b) results. However, if it is inserted after the first element, an ungrammatical string is formed (i.e. 7c). This shows that /*pomp-(e) benzin/* in the sense of ‘gas station’ is a CN as it has no internal inflection. Apart from internal inflection, this form can be tested for syntactic modifiability. If the adjective /*bozorg/* (“big”) is added to the end of the construct (as in 7d), the result is grammatical; but if it is put after the first element (as in 7e), the result is an ill-formed construct, which indicates the feature of no syntactic modifiability for this form. It leaves no doubt that /*pomp-(e) benzin/* meaning ‘gas station’ is a CN. There is a crucial point about this example that must be carefully noted. When this construct is used as an NP (i.e. 7f), it has a different meaning (i.e. ‘gas pump’). In this case, both (7c) and (7d) are grammatical because they are NPs, which of course have internal inflection and syntactic modifiability of their elements. Needless to say that they do not have the meaning of ‘gas station’ anymore. It seems that (7a) as a CN has derived from (7f) as an NP through compound lexicalization, whereby the process of center-switching has turned this form from an endocentric NP into an exocentric CN. As we have already seen in the case of exocentric N + A CNs, this process obliterates the syntactic features of an NP all together (i.e. internal inflection and syntactic modifiability) and transforms it into a CN at once. Under such circumstances, Ezafe appears to lose its grammatical function and turn into a kind of morphological element, i.e. a linker. What remains a question, however, is the reason for the optionality of Ezafe. This issue will be addressed later in the discussion.

It is also possible to find exocentric CNs which include obligatory Ezafe. (8a) is one such form.

- (8) a. *marg-e muš* (CN)
 death-EZF rat
 “raticide”
 b. *marg-e muš-hâ*
 death-EZF rat-PL
 “raticides”
 c. **marg-hâ-ye muš*
 d. *marg-e muš-e qavi*
 death-EZF rat-EZF strong
 “strong raticide”
 e. **marg-e qavi-ye muš*

- f. marg-e muš (NP)
 death-EZF rat
 “rat’s death”

Semantic idiosyncrasy verifies the status of this form as a CN. Lexical integrity confirms this fact as well. This construct has no internal inflection (as seen in 8b vs. 8c) and the first element is not syntactically modifiable (as seen in 8d vs. 8e). On account of these two criteria, there is no doubt that (8a) in the sense of ‘raticide’ is a CN. This form seems to have gone through a process similar to that of (7a). (8a) as an exocentric CN derives from (8f) as an endocentric NP through compound lexicalization, which involves center-switching. As a result, the original NP loses its syntactic properties, namely internal inflection and syntactic modifiability, and is transformed into a CN, where Ezafe also loses its grammatical function and turns into a linker. This example indicates that Ezafe is by no means a decisive criterion for NP status and CNs might appear with the element of Ezafe. The reason for the existence of Ezafe in CNs will be investigated in a forthcoming section of the present study.

As discussed earlier, the N + N pattern can be adopted to form endocentric CNs, which are realized in two different structures, namely head-final and head-initial constructs. Head-final constructs are by no means similar to NPs since they have their heads as the second element in the structure, whereas NPs are head-initial. Consequently, it is most unlikely that boundary cases can be found among head-final constructs. Head-initial constructs, on the other hand, are similar to NPs in that both of them have their heads as the first element in the structure. Due to this close resemblance, these constructs are expected to exhibit more complexity and challenge as far as the delicate issue of NP-CN demarcation is concerned. This is the area where boundary cases are most likely to appear. In the next part of the discussion, each one of the endocentric N + N CNs will be discussed in more detail.

Head-final CNs are those in which the second N plays the part of the head. One example of such words is given below:

- (9) a. ketâb xâne
 book house
 “library”
 b. ketâb xâne-hâ
 book house-PL
 “libraries”
 c. *ketâb- hâ xâne
 d. *ketâb-e târixi xâne (/târixi/ “historical”)

On account of the fact that a noun can only function as a post-modifier in an NP, but not a pre-modifier, the N + N construct can be an NP only if it is head-initial, i.e. N (head) + N (post-modifier). Consequently, the above-mentioned head-final construct can never have a corresponding phrase in syntax and thus it is not likely to be the result of lexicalization. It appears to be based on a noun compounding pattern that is morphological, rather than syntactic. In view of the arguments above, there is virtually no doubt that (9a) must be classed as CN. What confirms this conclusion further is the fact that this form lacks Ezafe, which is an indispensable feature of NPs. The criteria of internal inflection and syntactic modifiability obviously show that this construct is a CN (see 9c and 9d above). It should be noted that, on the whole, head-final N + N is a highly productive pattern for noun compounding in Persian.

In head-initial CNs, as the name suggests, the head appears at the onset of the construct. These head-initial forms correspond to N (head) + N (post-modifier) in NPs. A good illustration of such CNs is (10a) below:

- (10) a. surat hesâb (CN)
 statement account
 “financial statement”
 b. surat hesâb-hâ
 statement account-PL
 “financial statements”
 c. *surat-hâ-ye hesâb
 d. surat hesâb-e mâhiyâne
 statement account-EZF monthly
 “monthly financial statement”
 e. *surat-e mâhiyâne-ye hesâb
 f. surat-e hesâb (NP)
 statement-EZF account
 “the statement of an account”

As can be seen in this example, it does not have Ezafe. In this construct, internal inflection cannot be observed (10b vs. 10c). Also, this form does not have the feature of syntactic modifiability (10d vs. 10e). On the basis of the foregoing argument, it is evident that this form is a CN. It was derived from (10f) as an NP via the process of pure lexicalization involving complete syntactic erosion, in which all the three syntactic properties of Ezafe, internal inflection and syntactic modifiability vanish all together, transforming the NP into a CN.

3. N + N Boundary Cases

Not all head-initial constructs are as straightforward as the above example. As a case in point, (11a) below can be reviewed:

- (11) a. modir(-e) kol
 manager-EZF general
 “general manager”
 b. modir-hâ-ye kol (or modir-ân-e kol)
 manager-PL-EZF general manager-PL-EZF general
 “general managers”
 c. modir(-e) kol-hâ
 manager-EZF general-PL
 d. modir(-e) kol-e mehrabân
 manager-EZF general-EZF kind
 “a kind general manager”
 e. *modir-e mehrabân-e kol
 f. modir-e kol (NP)
 manager-EZF general
 “general manager”

It has optional *Ezafe* and may be equally used in both forms by Persian speakers. Moreover, this construct enjoys the feature of double inflection and plural affix might be added in two positions without any alteration in meaning (see 11b and 11c). The presence of *Ezafe* and the potential for internal inflection makes this form more similar to an NP. However, syntactic modifiability test in this case demonstrates that this form behaves more like a CN in spite of internal inflection and *Ezafe* inside it. If the adjective /*mehrabân*/ (“kind”) is used to expand /*modir*/ as the first N in the construct, which would be the head if the construct were an NP, the result is ill-formed (see 11e). The only acceptable form is (11d). It is now clear that this construct is a boundary case which shows contradictory results when different NP-CN criteria are applied. The evidence discussed above indicates that this construct was derived from the NP through the process of pure lexicalization which involves partial syntactic erosion. As a result of this process, syntactic modifiability disappears entirely, but *Ezafe* and internal inflection become optional, leading to the characteristic of double inflection in the construct.

There are also head-initial constructs which contain obligatory *Ezafe*, contrary to the typical instances of CNs. One such form is (12a):

- (12) a. mahd-e kudak
 cradle-EZF child
 “kindergarten”
 b. mahd-hâ-ye kudak
 cradle-PL-EZF child
 “kindergartens”
 c. mahd-e kudak-hâ
 cradle-EZF child-PL
 “kindergartens”
 d. mahd-e kudak-e bozorg
 cradle-EZF child-EZF large
 “a large kindergarten”
 e. *mahd-e bozorg-e kudak
 f. mahd-e kudak (NP)
 cradle-EZF child
 “child’s cradle”

It has *Ezafe* as an indispensable element without which this form can never be used. Semantically speaking, this form must be listed in the lexicon as it refers to a certain childcare facility and thus it enjoys semantic idiosyncrasy. This fact proves it to be a CN. However, lexical integrity yields contradictory results. While this construct has double inflection (12b and c), its first element does not take independent modifiers (12d vs. 12e). It seems that this construct is the consequence of partial syntactic erosion in the process of pure lexicalization of the NP (12f), which gives rise to the retention of *Ezafe*, the emergence of double inflection and the loss of syntactic modifiability.

An important question should now be addressed about the odd behavior of *Ezafe* in CNs beyond usual expectations. As mentioned in the previous discussions, *Ezafe* has three different states in nominal constructs: obligatory *Ezafe*, optional *Ezafe* and total exclusion of *Ezafe*. It was explained that NPs have obligatory *Ezafe*. In lexicalization, *Ezafe* is expected to undergo syntactic erosion and disappear entirely. Therefore, the absence of *Ezafe* in a nominal form means that lexicalization has been completed and a CN has been formed, as in (10a). In nominal constructs without *Ezafe*, internal inflection and syntactic modifiability cannot be observed (e.g. 10b vs. 10c and 10d vs. 10e). This does not always happen, however. In several cases, *Ezafe* remains in the construct as an obligatory or optional element. It is often argued that phonological limitations prevent the omission of *Ezafe* (Shaghghi, 2012). For instance, the

occurrence of obligatory Ezafe in /toxm-e morq/ (/toxm/ “egg”, /morq/ “hen”; chicken egg) and /taxt-e xâb/ (/taxt/ “bed”, /xâb/ “sleep”; bed) is due to phonological considerations in Persian and these forms are impossible or at least difficult to pronounce, so speakers either keep Ezafe or omit one phoneme to make the construct pronounceable, as in [toxm-e morq] or [tox-morq] (Shaghghi, 2012). This is what actually happens in the case of /pomp-(e) benzin/ discussed earlier in this section (see 7a). This form is either pronounced as [pomp-e benzin] with Ezafe inside it, or as [pom benzin] without Ezafe involving the elision of final [p] in [pomp]. Here, two points should be noted, however. First, forms like (4a) and (11a), which both have optional Ezafe, are not affected by any phonological processes (e.g. elision) if they occur without Ezafe. The optionality of Ezafe in such cases indicates that the occurrence of this element cannot merely result from phonological considerations in Persian. If this were the case, Ezafe would have to appear in the construct as an obligatory element or its omission would be accompanied by the elision of a phoneme in the construct. Second, no phonological limitations are observed in some constructs with obligatory Ezafe. (5a) does not basically need Ezafe to be pronounceable since it could be pronounced as [sagâbi], instead of [sag-e ʔâbi], through the process of re-syllabification. Therefore, Ezafe is not a requisite element in this construct and could easily be removed. In spite of this fact, Ezafe is obligatory in this form. Consequently, the occurrence of Ezafe in non-NP constructs must have another reason apart from phonological motivations. Upon reviewing the data, one can easily notice the fact that such forms all have internal inflection. The interesting phenomenon here is that the state of optionality or obligatoriness of Ezafe coincides with that of internal inflection. The potential for internal inflection in these constructs shows that Ezafe has not totally lost its grammatical function and does not solely play the role of a linker, but rather a syntactic element which makes internal inflection possible. In sum, the occurrence of Ezafe in nominal constructs is determined by two factors: syntactic function and phonological limitations. That is why the interpretation of Ezafe and its behavior in lexicalization proves to be rather difficult. If there are no phonological limitations, Ezafe is closely connected with internal inflection in that obligatory Ezafe coincides with obligatory internal inflection, as in (12a), optional Ezafe corresponds with optional internal inflection, as in (11a), and the exclusion of Ezafe rules out the possibility of internal inflection, as in (10a). However, in case of phonological motivations, the occurrence of Ezafe is obligatory regardless of the state of internal inflection and its omission is possible only if it is accompanied by a phonological process like elision.

IV. CONCLUSION

There are several different criteria for the demarcation of NPs and CNs. Phonological criterion, semantic criterion and listedness, and the principle of lexical integrity with the axes of internal inflection and syntactic modifiability are among the existing criteria. Some criteria, such as Ezafe, are peculiar to Persian and may not be applicable to other languages.

In Persian, there are several various patterns which create a large number of nominal constructs, most of which appear to be unambiguous in their nature as to whether they are NPs or CNs because they exhibit consistent results in the application of different NP-CN demarcation criteria. However, there are boundary cases among these nominal constructs, whose real nature is rather ambiguous as they react inconsistently to the existing criteria of demarcation. N + A and N + N are two nominal patterns which create boundary cases in Persian. It should be noted that these two patterns also produce both NPs and CNs. Most CNs of these patterns are derived from their respective NPs through the process of lexicalization, which involves syntactic erosion whereby an NP gradually loses its syntactic properties and is transformed into a CN. This syntactic erosion may occur in two forms: partial and complete. In complete syntactic erosion, all syntactic features of an NP disappear and a CN with a full pack of its characteristics is formed. In partial syntactic erosion, however, only some of NP syntactic properties vanish and the rest remain intact. Lexicalization also has two types: compound and pure. The former is accompanied by center-switching or category change, while the latter is not. In compound lexicalization with center-switching, the construct turns from an endocentric NP into an exocentric CN. The process of center-switching causes the construct to lose syntactic modifiability and internal inflection all together and, in the absence of any phonological obligations, Ezafe will be eliminated, too. Therefore, compound lexicalization entails complete syntactic erosion, yet pure lexicalization may involve either complete or partial syntactic erosion. In pure lexicalization, the center of an NP will not be altered. In such cases, semantic idiosyncrasy and listedness may occur. This process leads to endocentric, often head-initial, constructs.

Boundary cases which are formed on the basis of the two aforementioned patterns seem to share certain properties. They are found where there is maximal likeness to NPs as far as syntactic properties are concerned. This close similarity stems from the process of pure lexicalization which involves partial syntactic erosion of NPs. This forms intermediate constructs on the NP-CN continuum between the two extreme points of NP and CN, which partly resemble either constructs while bearing differences with both of them at the same time. Persian data, as reviewed earlier in this study, demonstrate that boundary cases are endocentric, a feature which makes them rather more similar to NPs. In addition, these forms are all head-initial in the case of N + A and N + N constructs like their respective NPs. Furthermore, almost all boundary cases do not have the property of syntactic modifiability, like usual CNs. As to internal inflection, however, boundary cases demonstrate a different state, compared to ordinary CNs; they necessarily have either obligatory or optional internal inflection. In the former case, where internal inflection is obligatory like NPs, boundary cases appear to bear striking similarity to their corresponding NPs, with lack of syntactic modifiability being

the only distinction between them. In the latter case, where constructs display optionality of internal inflection, the interesting phenomenon of double inflection arises, where a construct has two parallel plural forms with the same meaning and function: one similar to an NP with internal inflection and the other resembling a CN with final inflection. Another characteristic of N + A and N + N boundary cases is that they necessarily have Ezafe, whether it is obligatory or optional. This feature seems to be closely related to the presence of internal inflection: Ezafe co-occurs with the capacity for internal inflection in N + A and N + N patterns.

Among all criteria, the property of syntactic modifiability under lexical integrity can function as the most effective formal criterion for NP-CN demarcation since NPs seem to lose this characteristic in the early stages of lexicalization. Accordingly, almost all non-NP constructs share the feature of no syntactic modifiability. The test of internal inflection (pluralizability) in Persian bears only relative results since it works one-way. This means that if a form has no internal inflection, it is necessarily a CN, but no definite conclusion can be drawn if a form has internal inflection: a form with internal inflection might be either an NP or an intermediate construct on the NP-CN continuum. Ezafe has a similar status. Its absence from the construct is synonymous with the compound status of that construct, yet its presence does not prove anything as the construct might be an NP or an intermediate structure on the NP-CN continuum as well. It can even occur in a CN on account of phonological obligations. Therefore, the existence of Ezafe in a construct does not rule out the possibility of a CN at all.

REFERENCES

- [1] Afrashi, A. (2010). *S âxt-e zab ân-e f ârsi* (The structure of Persian language). Tehran: SAMT publications.
- [2] Anderson, S. R. (1988). Morphological theory, in F. Newmeyer (ed.), *Linguistics: the Cambridge survey* (vol. 1), Linguistic Theory: Foundations. Cambridge: CUP.
- [3] Anderson, S. R. (1992). *A-mrphous morphology*. Cambridge: Cambridge University Press.
- [4] Aronoff, M. (1976) *Word formation in generative grammar*. Cambridge, Mass.: MIT press.
- [5] Bateni, M. (1994). *Neg âhi t âze be dastur-e zab ân* (A new perspective on grammar). Tehran: Agah Publications Institute.
- [6] Bauer, L. (1983) *English word-formation*. Cambridge: CUP.
- [7] Blom, Corrien. (2005) The demarcation of morphology and syntax: A diachronic perspective on particle verbs, in Wolfgang, et al. (Eds). (2005) *Morphology and its demarcations*. Amsterdam: John Benjamins Publishing Co.
- [8] Bloomfield, L. (1933) *Language*. New York: Holt.
- [9] Booij, Geert. (2005). *The grammar of words: An introduction to Linguistic Morphology*. Oxford: Oxford University Press.
- [10] Bresnan, Joan & Sam Mchombo. (1995). The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory*, 13, 181–254.
- [11] Chomsky, N. (1957) *Syntactic structures*. The Hague: Mouton.
- [12] Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass.: MIT press.
- [13] Chomsky, N. (1970). Remarks on nominalization, in Jacobs and Rosenbaum (Eds) (1970) *Readings in English transformational grammar*. Waltham, Mass.: Blaisdell.
- [14] Chomsky, N. (1975) *The logical structure of linguistic theory*. New York: Plenum.
- [15] Chomsky, N. (1980) *Rules and representations*. New York: Columbia University Press.
- [16] Chomsky, N. (1981) *Lectures on government and binding*. Dordrecht: Foris.
- [17] Chomsky, N. (1986) *Knowledge of language, its nature, origin, and use*. New York: Praeger.
- [18] Dabir-Moghaddam, M. (1999). *Zabân šen âsi-ye nazari: peidâyeš va takvin-e dastur-e zâyeshi* (Theoretical linguistics: Emergence and development of generative grammar). Tehran: Sokhan Publications.
- [19] Di Sciullo, A. M., and Williams, E. (1987). *On defining the word*. Cambridge, Mass.: MIT Press.
- [20] Fodor, J. (1983) *The modularity of mind*. Cambridge, Mass.: MIT Press.
- [21] Fodor, J. (1989) Why should the mind be modular. In A. George (Ed.), *Reflections on Chomsky*, pp. 1-22, Blackwell.
- [22] Haspelmath, M. (2002). *Understanding morphology*. London: Arnold.
- [23] Katamba, Francis. (1993). *Morphology*. London: Macmillan Press LTD.
- [24] Lapointe, Stephen (1980). *A theory of grammatical agreement*. Ph.D. dissertation, University of Massachusetts, Amherst (published 1985, New York & London: Garland).
- [25] Marchand, H. (1969). *The categories and types of present-day English word formation*, 2nd edn. Munich: Beck.
- [26] Meshkatod Dini, M. (2002). *Towsif va ?âmuzeš-e zab ân-e farsi* (Introduction to aspects of the Persian language: The history, development of skills and structures). Mashhad: Ferdowsi University Press.
- [27] Newmeyer, F. (1983). *Grammatical theory: Its limits and its possibilities*. Chicago: University of Chicago Press.
- [28] Rohrer, C. (1974). Some problems of wordformation, in Rohrer C. and Ruwet, N. (Eds.) *Actes du colloque Franco-allemand de grammaire transformationnelle II*. Tübingen: Niemeyer.
- [29] Shaghghi, V. (2012). *Mab âni-ye sarf* (An introduction to morphology). Tehran: SAMT publications.
- [30] Williams, E. (1981). On the notions 'lexically related' and 'head of a word', *Linguistic Inquiry* 12: 245–74.

Shahram Peyman was born in Abadan, Iran in 1977. He holds a B.A. in English translation from Islamic Azad University, Abadan Branch, Abadan, Iran, and an M.A. in general linguistics from Ferdowsi University, Mashad, Iran. He is currently a PhD student majoring in general linguistics at, Isfahan, Iran. Since 2001, he has been lecturing in a number of universities in Iran, including Islamic Azad University, Abadan Branch, and AIT (Abadan Institute of Technology). His research interests include Persian syntax and morphology.

Vali Rezaei has a Ph.D in Linguistics from the University of Isfahan, Iran. He is an associate professor of Linguistics at the University of Isfahan. His research interests include syntax and typology. He is the primary author of *Fundamentals of Linguistic Typology* (in Persian) published by Shahid Beheshti University Press, Tehran (2016). He has published many articles in Iranian and international journals.