Study of Noun Phrase in Urdu

Aasim Ali
Punjab University College of Information Technology (PUCIT)University of The Punjab,Lahore, Pakistan
aasim.ali@pucit.edu.pk

Sarmad Hussain
Centre for Language Engineering Al-Khawarizmi Institute of Computer Sciences, UET Lahore
sarmad.hussain@kics.edu.pk

Kamran Malik
Institute of Business & IT, University of Punjab, Lahore, PK
mkamranmalik@gmail.com

Shahid Siddiq
National University of Computer & Emerging Sciences, Lahore Campus
shahid.siddiq@nu.edu.pk

Abstract

This paper presents discussion of internal structure of Noun Phrase (NP) in Urdu. It also proposes a computational grammar using LFG (Lexical Functional Grammar) formalism.

1. Introduction

As in other languages, Noun Phrase (اسمی ترکیب, NP) in Urdu also ranges from a single word to (theoretically) an infinitely long construction comprising other phrases as its constituents. NP may contain a number of word classes and phrases including pronouns, adjectives, nouns and quantifiers. It is important to find out the exact rules which govern how NP may be formed in Urdu and when these constructions are licensed. This is not only important to understand the grammar of Urdu but also essential for developing the computational models of the language. Case Phrase (KP[1]) in Urdu is formed by simply an NP or NP followed by Case Marker, cm). KP (and thus NP) forms fundamental argument for other constituent structures, including a Sentence (S) and Verb Phrase (VP). The variety in which NP may be constructed in addition to its frequent usage in forming or parsing Urdu grammar makes it an essential topic of further investigation.

Various aspects of construction of NP in Urdu have been addressed by many grammar books of Urdu. The current work looks at these analyses in conjunction with the requirements for Urdu grammar from the corpus of Urdu text to assess and extend the work. Analysis of different authors is discussed and eventually a formal set of rules for analyzing NP of Urdu is developed. Part of the published material on Urdu grammar is available only in Urdu hence the translation of the terminology used in such text is also presented in this paper. However, when exact English translation is not found in the literature, then most conceivable translation (as apparent from its definition and illustration given in that text) is provided instead of mere transliteration, so that the reader may realize the underlying concept and meaning. Lexical Functional Grammar (LFG) framework is used to realize these rules. Other formalisms can also be used, and the choice of LFG is arbitrary.

Next section visits the literature; the subsequent section shows the analysis and topic-wise segments of the computational grammar presented in this paper. Then, there is a comparative analysis of the literature and current work.

2. Literature Review

This section starts with identifying the scope of work, then after reviewing elementary

[1] Since CP is more generally known as abbreviation for other term Complementizer Phrase, hence KP is used here as abbreviation for Case Phrase, to avoid confusion. [14] has also used KP to abbreviate Case Phrase.

constituents of NP, the phrasal constituents are looked at, in this section of the paper.

2.1. Scope of this paper

This paper studies NP construction for the purpose of computation grammar for Urdu. The NP can contain clauses, other phrases, and elementary items as its constituent. Some constituents (phrases and clauses) of NP also have the ability of taking NP as their constituent (relative clause, postpositional phrase, and case phrase etc.); while others always appear at a subordinating position of NP (e.g. noun, pronoun, numeral, non-verbal adjective). This paper focuses on such subordinating items only, because each of other constructions requires its separate study and analysis. Therefore, genitive pronoun and verbal adjectives are also not included in this paper.

Handling of the coordinate conjunction of NP and adjectives is also included in the current grammar.

2.2. Elementary Items

This section reviews the literature available on elementary constituents of the NP.

2.2.1. The Noun (اسم)

Oxford English Dictionary [3] describes Noun as "A word used as the name or designation of a person, place, or thing; the class or category of such words." WordNet [1] defines noun semantically as "a content word that can be used to refer to a person, place, thing, quality, or action", or functionally as "the word class that can serve as the subject or object of a verb, the object of a preposition, or in apposition." Nouns are generally divided into many sub-classes, which include common vs. proper nouns, and mass vs. countable nouns (e.g. see [Jurafsky]). Further sub-classes may also be defined based on syntactic, grammatical or semantic roles of different nouns. Urdu grammarians have also sub-classified the nouns, which are discussed below.

Haq [6] divides the nouns in two main classes, proper (اسم خاص) and common (اسم عام). Proper nouns are further divided into four sub-classes, title (خطاب), attributive name (لقب), alias (عرف) and nom-de-plume (تخلص), and common nouns are sub-classified as state (کیفیت), collective (جمع), locative (ظرف) and instrumental (آلہ) nouns. Haq further divides the locative noun into nouns which represent a "location" in space (ظرفِ مکاں) and time (ظرفِ زماں).

Javed [5] divides the nouns flatly into five types, common (عام), proper (خاص), collective (جمع), abstract (مجرد) and mass (or non-countable, غیر شماری, also referred to as material, مادہ) nouns. Platts [9] does not explicitly discuss the types, but divides nouns into five classes based on morphological behavior. He lists abstract nouns, nouns of agency or attributes or possession or appellations, nouns of place, diminutive nouns and compounds.

Siddiqui [8] presents the most comprehensive classification of nouns. He divides the classification on the basis of structure (ساخت), nature (نوعیّت) and other. The classification on the basis of structure (ساخت) is as follows:

a. Primitive (neither derived nor derivable) Noun (اسم جامد), which is sub-classified as proper (خاص) and common (عام) nouns. The proper nouns are further sub-categorized as title (خطاب), Appellation or Attributive Name (لقب), Alias (عرف), nom-de-plume (تخلص) and kinship (کُنیّت). The common nouns are sub-categorized as state (کیفیت) and collective (جمع) nouns.

b. Verbal Noun (اسم مصدر), which is the infinitive and gerund form of verb, and generally used as noun. The derivational noun below is derived from verbal noun.

c. Derivational Noun (اسم مشتق), which is sub-classified into subject (فاعل), object (مفعول), present (حالیہ), deverbal (حاصل مصدر), wage (معاوضہ, name of the payment derived from the name of the work for which the payment is due), locative (ظرف, includes locative in "time" ظرفِ زماں and in "space" ظرفِ مکاں), and instrumental (آلہ) nouns.

The classification of nouns according to nature (نوعیّت) is as substantive (موصوف), quality (صفت) nouns and pronouns (ضمیر). Finally, Siddiqui lists other classes of nouns as sound (صوت), indefinite (کنایہ), relative (موصول), interrogative (استفہام), numeral (عدد), comparative (تفضیل) and exaggerative (مبالغہ) nouns.

[13] and [9] have not hierarchical categorization of noun in this way, they have studied the phenomenon of their morphological construction.

[8] has shown some categories of noun without embedding them into the hierarchy of main types, viz. sound noun (اسم صوت, names for sounds of animals or machines etc.), indefinite pronoun (اسم کنایہ, the translation is referred from [9] on the basis of description and illustrations of indefinite pronouns (translated by Platts as اسمائے تنیکر) given therein, includes words: کوئی (any) and کچھ (some)), relative pronoun (اسم موصول), interrogative pronoun (اسم استفہام), cardinal (اسم عدد), comparative adjective (اسم تفضیل), intensity adjective (اسم مبالغہ), and demonstrative pronoun (اسم اشارہ). Cardinal are discussed in numeral phrase (2.3.1 below); indefinite pronoun, interrogative pronoun, and demonstrative pronoun are discussed in pronouns (2.2.2 below); and comparative adjective and intensity adjective are discussed under adjective phrase (2.3.2 below).

As can be seen, different authors have done different classifications. These classifications are based on morpho-phonological, syntactic and semantic behavior of nouns.

Proper noun and common noun are the most common set (or subset) of types (noted in all the referenced texts). Urdu has no indication in the script to identify the proper noun (as opposed to the capitalization in English), but the grammatical distinction exists. Proper noun does not take plural declension while common noun does, in general. There are special uses of proper noun when it is not behaving typically:

[6] وہ شخص اپنے وقت کا [حاتم] ہے۔ "That person is *Haatam* of his days."
[6] وہ [رستم] ہند ہے۔ "He is *Rustam* of Hind."
[5] آج کل [حاتم] پیدا نہیں ہوتے۔ "These days *Haatams* are not borne"
[6] and [8] have also noted this exceptional usage of proper noun.

In the above mentioned usage, proper noun is used to refer to some property, state, or feature in the context, therefore [6] has called it adjective. [5] has not given it any specific name however he writes that whenever a proper noun is used in metaphoric behavior, it starts accepting *plural* as does the common noun. [8] has called such an instance of proper noun as adjective (because it is signifying some attribute of the entity owning this proper noun), for its being the sub-type of generalized noun (اسم نکرہ, a proper noun made common through تنکیر, Arabic word for *generalization*, so it can be termed generalized noun, henceforth), justifies it by distinguishing it from regular common noun.

The Infinitive form of Verb in Urdu mostly behaves as noun entry (and sometimes as a verb to show the aspect of necessity. [9] has used both labels (Infinitive and Verbal Noun) to refer to this type. [9] has also mentioned that Gerund in Urdu also has the same shape (as of Infinitive) and behaves as Verbal Noun.

[5] has stated Non-Count Noun and its synonym اسم مادہ, which may be translated as Material Noun or Mass Noun, to be the same concept. [12] has defined Material Noun as something that is used as ingredient of other product (made of/from the former), for example:

"This cake is made with lot of [eggs]. A bracelet of [solid gold]. A man of [steel]."

The first example indicating "*eggs*" as Material Noun is apparently contradicting with the definition of Non-Count Noun, but the other side of the picture is that the example of گیہوں (wheat) as Non-Count Noun seems apparently uncountable because of the size of each grain in proportion to their number, which somehow justifies the "*eggs*" especially in a bulk quantity (or may be when the number is not known). [5] is the only grammar writer who has noted this kind of the Noun. An important property of this type of noun is that they do not accept Numeral Adjective; rather they are measured in quantity so they accept Quantifying Adjective (a.k.a. Quantifier). Nevertheless, this kind is very important and useful to separate from other nouns especially when focusing on computational aspects for the sake of agreement with context (in terms of selection of one of the two, Numeral and Quantifier).

However, it is also observed that " دو پیالی چائے" (two cups of tea) is shortened to "دو چائے" (two teas). Here the measurement unit is presumed to be "پیالی", because of the nature of substance in question, and the mindset of the participants of conversation. In such uses of Numerals (with Non-Count Noun) context is very important, for example "دو چاول" (two rice) may refer to "دو پلیٹ چاول" (two plates of rice) or

"دو پیالے چاول" (two bowls of rice) depending on the measurement unit most frequently used between the participants of conversation.

2.2.2. The Pronoun (اسم ضمیر)

[9] has divided pronoun into demonstrative pronoun (صفت ضمیری), personal pronoun (ضمیر شخصی), relative pronoun (ضمیر موصولہ), correlative pronoun (جواب اسم موصول), interrogative pronoun (ضمیر استفہام), indefinite pronoun (ضمیر تنکیر), reflexive pronoun (ضمیر معکوس), reciprocal pronoun (ضمیر باہمی), and pronominal adjectives (صفات ضمیری). Only [9] has noted the pronoun type correlative, however, personal pronoun is used instead of correlative pronoun in modern Urdu [9].

[5] has classified pronouns into personal (with subclass honorific), reflexive, demonstrative, genitive (), common genitive, indefinite, interrogative, relative, adverbial (متعلق فعلی ضمیر, with subclasses, locative, temporal, and manner).

[6] has categorized pronouns into personal, relative, interrogative, demonstrative, and indefinite.

[8] has noted personal pronoun, relative pronoun, interrogative pronoun, demonstrative pronoun, indefinite pronoun, and pronominal adjectives to be the types of pronoun. [8] has divided the pronoun into demonstrative pronoun, personal pronoun, relative pronoun, interrogative pronoun, and indefinite pronoun. [8] has shown honorific pronoun (ضمیر تعظیمی) and reflexive pronoun to the sub-type of personal pronoun. [8] has also noted that the word جو (who) is the only relative pronoun in Urdu, and has other forms like جسے (to whom) is the oblique form.

[13] has shown demonstrative, personal, reflexive, interrogative, indefinite, relative, and repeated pronouns. The last type seems a morphological phenomenon of reduplication. However, the reduplications illustrated in the text are lexicalized (not generative).

2.3. Phrasal Items

This section reviews the literature available on phrasal constituents of the NP. Since, the genitive phrase (thus the genitive pronoun) is considered to be the KP hence, excluded from the discussion of NP.

2.3.1. The Adjectives (اسم صفت)

[6] has divided adjectives into personal adjective (صفت ذاتی), associative adjective (صفت نسبتی), cardinal and pronominal adjective (صفت عددی), quantifier (صفت مقداری), and demonstrative pronoun (صفت ضمیری).

[5] has categorized adjectives as personal adjective, cardinal and pronominal, quantifier, intensity adjective (صفت تشدیدی), demonstrative pronoun (صفت ضمیری), and comparative adjective (درجہ).

[8] has classified adjectives as personal adjective (صفت مشبہ), associative adjective, cardinal and pronominal adjective (also includes fraction (عدد کسری) as its subclass), and ordinal (صفت عددی).

[13] has shown interrogative pronoun, indefinite pronoun, adverbial pronoun (near, far, interrogative, and relative), repeated adjective, and comparative adjective to be the types of adjectives. [13] has also noted possessive adjective, reflexive possessive adjective, numbers, and participles used as adjective.

[9] has studied morphological behavior of Urdu adjectives.

[5] has shown that the adjective usually precedes the noun but for emphasis it may follow. In the following example the adjective is used at the end of the sentence:

جمیل تھکا ماندہ گھر پہنچا۔ (Jamil was tired when reached home.)
میں نے ایک غار دیکھا۔ بھیانک، تاریک، اور گہرا۔ (I saw a cave – scary, dark, and obscure)

The difference of order in the above examples is to emphasize on the specific attribute of the substantive.

[9] has noted the use of genitive construction (using genitive diacritic زیر اضافت), according to the rules of borrowed Persian Morphology (hence generally used with borrowed Persian words), where Adjective comes after the Head Noun in the Noun Phrase, e.g. مردِ نیک (a good man, literally "a man of good"). However, this construction is not frequently used in contemporary Urdu.

2.3.2. The Numerals

[9] has listed numerals apart from adjectives, and has subdivided them into cardinal, ordinal, collective numeral, distributive and multiplicative, numeral adverb, fractional number, currency (rakam).

[13] has also listed numerals apart from adjectives, and has categorized them as cardinal numbers, ordinal numbers, fractions, frequency (once, twice, etc.), multiplicative, repeated numbers, and numbers with oblique plurals.

[6], [5], and [8] have noted numerals as type of adjectives, hence, listed in the adjec-tive phrase (section 2.3.2) below.

2.4. Features and Agreement

The term لوازم اسم (Essentials of Noun) is used by Maulvi Abdul Haq [HQ1, p 14], to discuss the features of the constituents of the NP, and their agreement requirement to form a grammatically valid NP. Every noun (and some other constituents of NP, e.g. adjective) in Urdu has the Gender (جنس) feature, which needs to agree with the head of the NP being formed. When a constituent (e.g. whole number) does not have the gender restriction for attachment, it is said to have neutral gender which means it has set of all possible values for that feature in Urdu, i.e., masculine and feminine, as the value for its gender feature.

Productive Morphotactics have been observed in Urdu for the identification of the gender and number of a noun from its shape, but in reality, it is a lexical phenomenon, e.g. لڑکا (boy, example of marked masculine singular noun; the mark is ending vowel) vs. دعا (prayer, example of a noun having same mark but it is feminine singular), and لڑکی (girl, example of a marked feminine noun; the mark is ending vowel) vs. ہاتھی (elephant, example of a noun having same mark but it is masculine singular).

The feature of honor, in Urdu, overrides the feature of number: a singular starts requiring plural agreement due to increased honor, e.g. استاد آیا (The teacher came) vs. اساتذہ آئے (The teachers came) vs. استاد صاحب آئے (Respected teacher came). It also shows that the agreement is not restricted to NP and its constituents only, rather is witnessed across other phrases of the sentence, as well.

[11] has noted that explicit case (حالت) marking is useful for the establishment of the *semantic roles* of nouns (and pronouns) and their *syntactic relationship* to the verb. The freedom in phrase order in Urdu language is due to explicit case marking. For example:

لڑکی [نے] دروازے [کو] دیکھا۔ gives the same meaning as
دروازے [کو] لڑکی [نے] دیکھا۔ (The girl looked at the door.)

In the above example, "نے" marks that "لڑکی" is the agent (or the Subject, doing the act of *see*ing) and "کو" marks the preceding noun "دروازہ" to be the patient (or the Object, the thing being seen). So the case is a tool for marking relationships between dependents and the head [11].

Shape (form) of the gender marked noun also changes on the basis of attachment of Case Marker [13]:

(a) Nominative (فاعلی) form: Only for Nominative Case
(b) Vocative (ندائی) form: Only for Vocative Case. This form does not exist for Verbal Nouns [9].
(c) Oblique (غیرندائی) form: For all other Cases (requiring separate word, case marker, to follow)

Usually a countable noun allows plural inflection and can follow a cardinal or pronominal adjective, whereas a non-count noun allows pronominal attachment of quantifier. This constraint of attachment cannot be implemented without having a feature to indicate whether the noun is a count or non-count.

3. Proposed Grammar with Analysis

The proposed grammar is written using the notation of Lexical Functional Grammar. It is incremented with discussion on every addition.

3.1. The Noun

The computational analysis of the Noun portion of the SubstantivePhrase (NP consisting substantive noun only) has shown that theoretically infinite substantive nouns can occur in an NP.

```
SubstantivePhrase →
      [SubstantiveNoun]*
      SubstantiveNoun
```

The SubstantiveNoun is a POS (Part of Speech) that means Substantive Noun, this term is used after [8] and [9]. This word is more precise and specific than the term Noun that has broader sense. A substantive noun refers to the generic concept that includes common noun and proper noun etc. Adding the features to this grammar rule gives:

```
SubstantivePhrase→
       [SubstantiveNoun]*
       SubstantiveNoun: ↑Gender=↓Gender,
      ↑Number=↓Number, ↑Case=↓Case,
      ↑Count=↓Count, ↑Honor=↓Honor;
```

Since the unification works for the agreement of feature values of the current item with its composing parent hence only last noun in the string of nouns is coded for agreement, considering the last noun to be the head of the NP, most of the times, in Urdu.

3.2. Pronoun

As the genitive phrase is not in the scope of this paper therefore it is not modeled herein.

```
NounPhrase →
[
 (Demonstrative: ↑Gender=↓Gender,
  ↑Number=↓Number, ↑Case=↓Case,
  ↑Count=↓Count, ↑Honor=↓Honor;)
 SubstantivePhrase:  ↑Gender=↓Gender,
   ↑Number=↓Number, ↑Case=↓Case,
   ↑Count=↓Count, ↑Honor=↓Honor;
]
  |
[
 Pronoun: ↑Gender=↓Gender,
   ↑Number=↓Number, ↑Case=↓Case,
   ↑Count=↓Count, ↑Honor=↓Honor;
]
```

In the above rule, Demonstrative refers to the demonstrative pronoun, and parentheses around this POS show its being optional. The SubstantivePhrase, so far, is same as coded above. Pronoun is the POS assigned to all pronoun types except the demonstrative pronoun and genitive pronoun. It is important to notice that the Pronoun construction is prevented from Demonstrative attachment.

3.3. Adjective Phrase (ترکیبِ توصیفی)

The Adjective Phrase may contain multiple adjectival items.

```
AdjectivalPhrase →
       [Adjectival: ↑Gender=↓Gender,
        ↑Number=↓Number, ↑Case=↓Case,
        ↑Count=↓Count, ↑Honor=↓Honor;]*
       Adjectival: ↑Gender=↓Gender,
        ↑Number=↓Number, ↑Case=↓Case,
        ↑Count=↓Count, ↑Honor=↓Honor;
```

SubstantivePhrase must be modified to incorporate the AdjectivalPhrase (Adjective Phrase):

```
SubstantivePhrase →
       (
        AdjectivalPhrase: ↑Gender=↓Gender,
         ↑Number=↓Number, ↑Case=↓Case,
         ↑Count=↓Count, ↑Honor=↓Honor;
       )
       [SubstantiveNoun]*
       SubstantiveNoun: ↑Gender=↓Gender,
         ↑Number=↓Number, ↑Case=↓Case,
         ↑Count=↓Count, ↑Honor=↓Honor;
```

The AdjectivalPhrase is kept optional. The subsections in this section analyze and build the grammar for incorporation of the constituents of adjective phrase.

3.4. The Numeral Phrase

General numerals are complex and of recursive nature so it is better to devise a component/tool that assists the grammar application in this regard. For example:

تین سوپندرہ لاکھ ساٹھ ہزار چار سو تیرہ۔ (three hundred and fifteen lakh sixty thousand four hundred and thirteen)

3.4.1. Quantifier

Quantifiers occur before typical adjectives.

```
AdjectivalPhrase→... |
      [
       (Quantifier: ↑Gender=↓Gender,
         ↑Number=↓Number, ↑Case=↓Case,
         ↑Count=↓Count, ↑Honor=↓Honor;)
       AdjectivalPhrase: ↑Gender=↓Gender,
         ↑Number=↓Number, ↑Case=↓Case,
         ↑Honor=↓Honor;
      ]
```

3.4.2. Cardinal

Cardinal is the whole number, and may introduce extra complexity to the grammar if generated as ordinary production rule. So it was analyzed that to keep the grammar relevant to the

grammatical aspects. This idea entails that there exists a small program invoked from the grammar to generate or validate the complicated formation of the (theoretically) infinitely long whole number, referred to as CompositeCardinal here.

```
NumeralPhrase→
      [
       CompositeCardinal: ↑Gender=↓Gender,
        ↑Number=↓Number, ↑Case=↓Case,
        ↑Count=↓Count, ↑Honor=↓Honor;
      ]
```

Thus the AdjectivalPhrase will become:

```
AdjectivalPhrase→... |
      [
       (NumeralPhrase: ↑Gender=↓Gender,
         ↑Number=↓Number, ↑Case=↓Case,
         ↑Count=↓Count, ↑Honor=↓Honor;)
       [SubstantiveNoun]*
       SubstantiveNoun: ↑Gender=↓Gender,
         ↑Number=↓Number, ↑Case=↓Case,
         ↑Count=↓Count, ↑Honor=↓Honor;
      ]
```

In the above rule, SubstantivePhrase is extended (the ellipses are just shown here to indicate that previous rule is extended, not replaced) to incorporate NumeralPhrase is kept optional, when it occurs it contains nothing but a CompositeCardinal, till this stage. Feature agreement/unification ensures that value of the cardinal must agree with the number feature of the head of the SubstantivePhrase (which is the parent production of the NumeralPhrase). The CompositeCardinal constructs or validates the construction of the number which is a valid string of Cardinals where Cardinal is the POS assigned to such words.

3.4.3. Ordinals

Occurrence pattern of ordinal is same as that of demonstrative pronoun, so the NounPhrase must incorporate it at par with Demonstrative:

```
NounPhrase →
[
 (
  [
   Demonstrative: ↑Gender=↓Gender,
    ↑Number=↓Number, ↑Case=↓Case,
    ↑Count=↓Count, ↑Honor=↓Honor;
  ]
     |
  [
   Ordinal: ↑Gender=↓Gender,
    ↑Number=↓Number, ↑Case=↓Case,
    ↑Count=↓Count, ↑Honor=↓Honor;
  ]
 )
 SubstantivePhrase:  ↑Gender=↓Gender,
   ↑Number=↓Number, ↑Case=↓Case,
   ↑Count=↓Count, ↑Honor=↓Honor;
] ...
```

This is changed NounPhrase rule (not added). Here Ordinal is the POS for ordinal words in the lexicon. It must be noticed that whole construction before the SubstantivePhrase in this rule is optional.

3.4.4. Multiplicative

Multiplicative also cause similar extension the grammar so far.

```
NumeralPhrase→    ... |
[
 Multiplicative: ↑Gender=↓Gender,
  ↑Number=↓Number, ↑Case=↓Case,
  ↑Count=↓Count, ↑Honor=↓Honor;
]
```

Multiplicative is the POS for such words in the lexicon; and the ellipses are just shown here to indicate that previous rule is extended, not replaced.

3.4.5. Fractions

Urdu has two types of fractions, viz. Suffix Fractions (a Cardinal may precede, تین چوتھائی (three-fourth)), and Prefix Fractions (a Unit Cardinal always follows, ڈیڑھ ہزار (one and a half thousand)).

```
NumeralPhrase→    ... |
[
 (Cardinal)
 FractionalSuffix: ↑Gender=↓Gender,
  ↑Number=↓Number, ↑Case=↓Case,
  ↑Count=↓Count, ↑Honor=↓Honor;
] | [
 (FractionalPrefix)
 UnitCardinal: ↑Gender=↓Gender,
   ↑Number=↓Number, ↑Case=↓Case,
   ↑Count=↓Count, ↑Honor=↓Honor;
]
```

Cardinal, FractionalSuffix, FractionalPrefix, and UnitCardinal are the POS for the words of these categories; and the ellipses are just shown here to indicate that previous rule is extended, not replaced.

3.5. POS Modifying Particles

There a few morphemes in Urdu that are handled syntactically in this work because of their orthographic separation from previous token.

3.5.1. vaalaa (والا) Construction

The word والا[2] (say, vaalaa-morpheme) has an importance while working on computational grammar of Urdu; it adds complexity to the analysis of NP. There are some suffixes shown in [8] that can be used to make noun from an existing verb, of which vaalaa-morpheme is the most open to be used as suffix for this purpose. In fact, it forms Adjective Phrase, and occurs before the head noun of the Noun Phrase. Just like any other adjective, vaalaa-construction, also becomes a noun when the head noun is absent in the NP, for example:

پاگل آدمی بولا- (The mad man spoke.) is same as پاگل بولا- (The mad man spoke.); here dropping of the head of the NP causes the pertinent Adjective Phrase to be promoted to NP. Similarly,

دودھ والا آدمی بولا- (The milk-man spoke.) is same as دودھ والا بولا- (The milk-man spoke.); here, again, dropping of the head is inducing promotion of the Adjective Phrase to NP. However, the trend of dropping the head noun in vaalaa-construction is frequent.

The noun type, Generalized Noun to be the sub-type of Adjective, stated by [8], gives the linguistic evidence of this phenomenon. Flexibility of vaalaa-morpheme to allow any complex construction be preceded, adds complication in its computational analysis. It can be used with verbs and nouns. It can be arbitrarily long and complex as:

صبح کی جانے والی بات والے آدمی والی کتاب والے رنگ والی چوکھٹے والی تصویروالی لڑکی والارومال

(The handkerchief belonging to the girl in the picture with the frame whose color was same as the color of the book belonging to the man who was talked about in the morning conversation)

This construction seems ridiculous but is valid, grammatically. Therefore, keeping such cons-tructions generative via grammar seems logical.

Thus next grammar rule is accommodating the vaalaa-construction. For the purpose of this rule, a new Part of Speech "والاصرفیہ" (vaalaa-morpheme) is needed, which has only one main entry "والا", in all its forms (on the bases of gender, number, and case of the following word).

[2] The term, vaalaa-morpheme, introduced here, serves for both genders and both numbers in Urdu. However, surface form is Masc.Sg:والا, Fem.Sg: والی , Masc.Pl: والے, and Fem.Pl: والیاں.

```
vaalaaPhrase → NounPhrase
      vaalaa-morpheme: ↑Gender=↓Gender,
       ↑Number=↓Number, ↑Case=↓Case,
       ↑Count=↓Count, ↑Honor=↓Honor;
SubstantivePhrase →     ... |
      [vaalaPhrase]*
      (
       SubstantiveNoun: ↑Gender=↓Gender,
        ↑Number=↓Number,   ↑Case=↓Case,
        ↑Count=↓Count, ↑Honor=↓Honor;
      )
```

3.5.2. Numeral particles

There are some numeral morphemes that change the grammatical behavior of main word. These are referred to as Numeral Particles collectively.

The ellipses are just shown here to indicate that previous rule is extended, not replaced.

```
NumeralPhrase→    ... |
      [
       CompositeCardinal
       MultiplicativePartical
      ]
NumeralPhrase→    ... |
      [
       CompositeCardinal
       FractionalPartical
       CompositeCardinal
      ]
```

4. Conclusion and Future Work

The literature review from native and foreign grammar writers gave a detailed understanding about the internal structure of the Noun Phrase in Urdu. It is a complex structure; still the work is in progress. The following items are yet to be included in this grammar: Intensity sub-phrase under the Adjectives, More about vaalaa Construction, Units and Rates, Combination of Numerals with Adjectives, Combination of Numerals with Nominal, Coordinate Conjunction of NP (آخری تین نیلے ڈبے اور دو سادہ گنّے), and Coordinate Conjunction of Adjectival Phrase (تین لمبوترے، چار گول، اور دو چوکور ٹکڑے).

5. References

[1] WordNet – Princeton University Cognitive Science Laboratory, 1998 http://wordnet.princeton.edu, sited April 2007

[2] Mariam Webster's Online Dictionary, Mariam Webster Incorporation, 2007 http://www.m-w.com/, sited April 2007

[3] Oxford English Dictionary, Oxford University Press, 2007 http://www.oed.com/, sited April 2007

[4] Ali, Aasim. Report on Urdu Computational Morphology, 2007. A report prepared as integral requirement of reading course on "Urdu Computational Morphology", National University of Computer and Emerging Sciences, Lahore.

[5] Javed, Ismat. "نئی اردو قواعد", 1981. Taraqqi Urdu Bureau, New Delhi.

[6] Haq, M. Abdul. "اردو صرف و نحو", 1987. Anjuman-e-Taraqqi Urdu (Hind).

[7] Dr Haq, M. Abdul. " قواعدِ اردو". Lahore Academy, Lahore.

[8] Siddiqui, Dr Abu-ul-Lais. "جامع القواعد", 1971. Markazi Urdu Board, Lahore.

[9] Platts, John T. "A Grammar of the Hindustani or Urdu Language", 1909. London.

[10] Butt, Miriam, et el. "A Grammar Writer's Cookbook", 1999. Center for the Study of Language and Information, Stanford, California.

[11] Butt, Miriam. "Theories of Case", 2004. Cambridge University Press.

[12] Quirk, Randolph, et el. "A Comprehensive Grammar of the English Language", 1985. Longman Group Limited, England.

[13] Schmidt, R L. "Urdu: An essential grammar", 1999, Routledge, London and New York.

[14] Butt, M. and King, T.H. "The Status of Case". In V. Dayal and A. Mahajan (eds.) Clause Structure in South Asian Languages. 2005. Berlin: Springer Verlag, 153–198.