



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No :

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	06 <sup>th</sup> November, 04	0.1	Created

POS ID: UPOS01

Part of Speech Name: case marker

Part of Speech String: cm

Lexical Entry Template:

<word>: cm, ^CASE = {GEN, LOC\_TAK, LOC\_PAR, LOC\_MAIN, INST, DAT, ACC} ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

Sample Lexical Entry:

کا:cm, ^CASE=GEN, ^ NUM = SG, ^ GEND= M, ^ FORM = NOM, ^ RESPECT = NORESPECT.

کے:cm, ^CASE=GEN, ^ NUM = SG, ^ GEND = M, ^ FORM = OBL, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

**Description:** Case markers show case of noun phrase. These case values help in determining grammatical function of NP in the sentence.

The description of the features is following:

CASE can be:

GEN Example: کا  
LOC\_TAK Example: تک  
LOC\_PAR Example: پر  
LOC\_MAIN Example: میں  
INST Example: سے  
DAT Example: کو  
ACC Example: کو

NUM (Number) can be:

SG (singular): Example: کا  
PL (plural): Example: کے  
It is only applicable to genitive case marker.

GEND (Gender) can be:

M (masculine): Example: کا  
F (feminine): Example: کی  
It is only applicable to genitive case marker.

FORM can be:

NOM (Nominative): Example: کا  
OBL (Oblique): Example: کے  
It is only applicable to genitive case marker.

RESPECT can be:

NORESPECT, FAMILIAR, USUAL, EXTRA



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

It is only applicable to genitive case marker.

CASE can be:

ERG (Nominative): Example: کا  
ACC (Oblique): Example: کے

**Examples:**

تک ، میں ، کا ، کو ، کی ، کے ، پر ، نے ، سے

**POS Status:** Active

**Reference:** Butt and King, "The Status of case"

**Related POS:** UPOS102, UPOS103

**Replaces:** -UPOS01

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the rule.

*Analysis 1:* case markers are headless words or clitics. They only provide case feature (CASE) to their parent phrase. Possible case values are predefined. These values are as follows.

نے	Ergative	ERG
کو	ACC	ACC
کو	Dative	DAT
سے	Instrumental	INS
کا ، کی ،	Genitive	GEN
کے		
تک ، میں ،	Locative	LOC
پر		

*Examples:*

لڑکے نے  
\*لڑکا نے  
اس نے  
\*وہ نے

[ [حامد نے] [شیر کو] [ [افریقہ کے] [جنگل میں] [بندوق سے] [مارا.] ] ]

Examples of its Lexical Entries are mentioned above

*Analysis 2:* An alternative analysis could be that these words are prepositions and subcategorize NP.

**Result:** We decided on Analysis 1.

**Future Work:** Case values for locative case markers can be further divided.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	06 <sup>th</sup> November, 04	0.1	Created

POS ID: UPOS02

Part of Speech Name: Postposition

Part of Speech String: p

Lexical Entry Template:

<word>:p ^Pred = 'logical form of word<OBJ>', PreAllowedCase = {Gen}.

Sample Lexical Entry:

سمیت:p, ^Pred = 'سمیت<OBJ>', \_PreAllowedCase = Gen.

بین:p, ^Pred = 'بن<OBJ>', \_PreAllowedCase = Gen.

تلیے:p, ^Pred = 'تلیے<OBJ>', \_PreAllowedCase = Gen.

Description:

PreAllowedCase can be:

Gen Example: بن

Inst Example: بغیر

It can also be NULL

Examples:

تلیے ، لیے ، بن ، بغیر ، سمیت



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**POS Status:** Active

**Reference:** John T. Platts, "*A Grammar of the Hindustani or Urdu Language*"

**Related POS:** UPOS101, UPOS103

**Replaces:** -UPOS02

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the rule.

<word>:p ^Pred = 'logical form of word<OBJ>', PreAllowedCase = {Gen,Inst}.

*Analysis 1:* Postpositions subcategorize for an object which is a noun phrase. Object must be in oblique form. Object can be in Nominative or Genitive case depending on the particular word. Some Examples are:

آسمان تلے (Postposition Allowing Nominative Case)

درخت کے تلے (Postposition Allowing Genitive Case)

Some of these are not purely postpositions rather they can act as prepositions. In that case, case of object cannot be Nominative. For Example:

بغیر سواری کے (Preposition Allowing Genitive Case)

Some analyzed words and their features are as follows.

Words	Post Position / Preposition
سمیت	Post Position
تلے	Post Position
لیے	Post Position
بن	Postposition / Preposition
بغیر	Postposition / Preposition

Pre allowed case can have two values {Gen, Inst}. It can also be NULL

**Result:** We decided on above analysis.

**Future Work:**



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	06 <sup>th</sup> November, 04	0.1	Created

POS ID: UPOS03

Part of Speech Name: Nominal Preposition

Part of Speech String: nomp

Lexical Entry Template:

word:nomp ^Pred = 'logical form of word', PreAllowedCase = {Gen,Inst.

Sample Lexical Entry:

باہر:nomp, ^Pred = 'باہر', \_PreAllowedCase = {Gen,Inst}.

نیچے:nomp, ^Pred = 'نیچے', \_PreAllowedCase = {Gen,Inst}.

Description:

Examples:

اندر، باہر، اوپر، نیچے، پیچھے، آگے، ساتھ، نزدیک، پاس

POS Status: Active

Reference:

Related POS: UPOS101, UPOS102

Replaces: -UPOS03

Reason: -

Replaced by: -

Reason: -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the rule.

*Analysis 1:* Nominal Postpositions can subcategorize for a noun phrase as object and occur independently. Object must be in oblique form. Object can be in Nominative, Genitive or Instrumental case depending on the particular word. Main distinction in Nominal Postposition and other postpositions is that phrase formed by Nominal Postposition can precede a case marker. Some Examples are: (Nominative Postposition Only)

وہ [باہر] گیا۔ (Nominative Postposition Only)

وہ [گھر کے باہر] گیا۔ (Nominative Postposition allowing Genitive Case)

وہ [[گھر کے باہر] تک] گیا۔ (Nominative Postposition allowing Genitive Case and having Case Marker)

Some analyzed words and their features are as follows.

Word	Allowed Case	With Genitive	With Instrumental
اندر	Instrumental, Genitive	سے، تک، کے، کا، کو، کی	کو، کا، تک، کے، کی
باہر	Instrumental, Genitive	سے، تک، کے، کا، کو، کی	کو، کا، تک، کے، کی
اوپر	Instrumental, Genitive	سے، تک، کے، کا، کو، کی	کو، کا، تک، کے، کی
نیچے	Instrumental, Genitive	سے، تک، کے، کا، کو، کی	کو، کا، تک، کے، کی
پیچھے	Instrumental, Genitive	سے، تک، کے، کا، کو، کی	کا، تک، کے، کی
آگے	Instrumental, Genitive	سے، تک، کے، کا، کو، کی	کو، کا، تک، کے، کی
ساتھ	Genitive	کا، کے، کی، سے	
نزدیک	Instrumental, Genitive	سے، تک، کے، کا، کو، کی	کا، کے، تک، کی
پاس	Instrumental, Genitive, Nominative	سے، تک، کے، کا، کی	

*Analysis 2:* An alternative analysis could be that these words are prepositions and their difference with postpositions is catered by adding some additional feature to it.

**Result:** We decided on Analysis 1.

**Future Work:** Some of these may not be purely postpositions rather they can act as prepositions. Some references give such examples from old Urdu, but it is not frequent. This feature of these words is not yet being catered but may be added later.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Tafseer Ahmed	6 <sup>th</sup> Nov, 04	0.1	Created
Sara Hussain	7 <sup>th</sup> Jan, 2005	4.1.0.0	Added SEM_TYPE feature with TEMPORAL and PLACE options

POS ID: UPOS104

Part of Speech Name: Noun

Part of Speech String: n

Lexical Entry Template:

<word>: n, ^ PRED = '<logical form of word>', ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ FORM = {NOM, OBL}, ^ CTYPE = {COUNT, MASS}, ^ NTYPE = {PROPER, COMMON}, ^ SEM\_TYPE = {TEMPORAL, PLACE}, ^ ANIM = {+, -}, ^ RESPECT = {NORESPLECT, FAMILIAR, USUAL, EXTRA}.

Sample Lexical Entry:

لڑکی : n, ^Pred = 'لڑکی', ^ GEND = F, ^ NUM = SG, ^ FORM = {NOM, OBL}, ^ CTYPE = COUNT, ^ NTYPE = COMMON, ^ ANIM = +, ^ RESPECT = {NORESPLECT, FAMILIAR, USUAL, EXTRA}.

لڑکیاں : n, ^Pred = 'لڑکی', ^ GEND = F, ^ NUM = PL, ^ FORM = NOM, ^ GEND = F, ^ CTYPE = COUNT, ^ NTYPE = COMMON, ^ ANIM = +, ^ RESPECT = {NORESPLECT, FAMILIAR, USUAL, EXTRA}.

لڑکیوں : n, ^Pred = 'لڑکی', ^ GEND = F, ^ NUM = PL, ^ FORM = OBL, ^ CTYPE = COUNT, ^ NTYPE = COMMON, ^ ANIM = +, ^ RESPECT = {NORESPLECT, FAMILIAR, USUAL, EXTRA}.

لڑکی : n, ^Pred = 'لڑکی', ^ GEND = F, ^ NUM = SG, ^ FORM = {NOM, OBL}, ^ CTYPE = COUNT, ^ NTYPE = COMMON, ^ ANIM = +, ^ RESPECT = {NORESPLECT, FAMILIAR, USUAL, EXTRA}.

آج : n, ^Pred = 'آج', ^ GEND = M, ^ NUM = SG, ^ FORM = {NOM, OBL}, ^ CTYPE = COUNT, ^ NTYPE = COMMON, ^ SEM\_TYPE = TEMPORAL, ^ ANIM = -, ^ RESPECT = NORESPECT.

گھر : n, ^Pred = 'گھر', ^ GEND = M, ^ NUM = SG, ^ FORM = {NOM, OBL}, ^ CTYPE = COUNT, ^ NTYPE = COMMON, ^ ANIM = NEG, ^ RESPECT = NORESPECT.

**Description:** A noun is the name of a person, place, thing, or idea. Whatever exists, we assume, can be named, and that name is a noun.

The description of the features is following:

NUM (Number) can be:

SG (singular): Example: لڑکی

PL (plural): Example: لڑکیاں

It cannot be null or both.

GEND (Gender) can be:

M (masculine): Example: لڑکا

F (feminine): Example: لڑکی

It cannot be null or both. For mass nouns, it is always singular.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

FORM can be:

NOM (Nominative): Example: لڑکیاں  
OBL (Oblique): Example: لڑکیوں  
It cannot be null or both.

CTYPE (Count Type) can be: COUNT

Example: کتاب  
MASS Example: پانی  
It cannot be null or both.

NTYPE (Noun Type) can be: COMMON

Example: لڑکا  
PROPER Example: احمد  
It cannot be null or both.

SEM\_TYPE (Semantic Type) can be: TEMPORAL

Example: آج  
PLACE Example: گھر  
It cannot be both but can be NULL.

ANIM (Animated) can be:

+ (Positive): Example: لڑکا  
- (Negative): Example: کتاب  
It cannot be null or both.

RESPECT can be:

NORESPECT, FAMILIAR, USUAL, EXTRA  
It cannot be null. It is used for agreement especially in Predicate Link agreement.

The detail of the feature values is discussed in analysis section.

**Examples:**

شاعری، کمپیوٹر، ڈائری، پینسل، بستہ، بوتل، کتاب، آدمی، گھوڑا، حامد، احمد  
(Temporal) آج، ہفتہ، صبح  
(Place) گھر، کلکتہ، لاڑکانہ



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**POS Status:** Active

**Reference:**

- [1] <http://webster.commnet.edu/grammar/nouns.htm>  
[2] "Urdu Sarf au Nahav", Maulvi Abdul Haq

**Related POS:**

**Replaces:** - UPOS04

**Reason:** -

**Replaced by:** -

**Reason:** -



**Center for Research in Urdu Language Processing**  
**National University of Computer and Emerging Sciences, Lahore Pakistan**

**Analysis:** Following is the in-depth analysis of the rule.

**Analysis:**

Number of a noun can be singular (SG) or plural (PL) or both. It is a mandatory feature. If any noun has both singular and plural values they will be written in the form of set, e.g. {SG, PL}. Same is for Gender feature.

In Urdu, noun can be of two forms. One is Nominative (NOM) i.e. *lerka* لڑکا and other is Oblique (OBL) i.e. *lerkey* لڑکے as singular. Nominative form is used when noun phrase is in Nominative case. When noun phrase is in some other case Oblique form is used along with case marker. For further details see POS Document of case markers and Grammar rule document for Case Phrase (KP).

Nouns can be of two types with respect to their physical properties: countable and mass nouns. This property helps in choosing right specifier for them. It is reflected as feature CTYPE. 'کتاب' is an example of Count Noun while 'پانی' is an example of Mass Noun.

With respect to their use, two major categories of nouns are common noun and proper noun. Proper Noun is a name of particular person, place or thing whereas common noun is a noun which can be used for generic type like man, horse, tree, book. These types are reflected in feature NTYPE. These are useful because on their basis some classes of specifiers and modifiers can be restricted to some typical type.

A noun can be animated or unanimated. It can help in deciding correct case marker for noun in some cases. For example, when a noun occurs at object position, accusative case is used for animated nouns and nominative case is used for unanimated nouns. e.g. اس نے کرسی اٹھائی، اس نے بچے کو اٹھایا.

In Urdu there is a rule to use plural agreements for singular nouns if one wants to give respect. Therefore a feature of respect is added which will have RESPECT value for words like 'چچا', 'والدہ' etc. Such words will have USUAL respect. For words like محمد ﷺ respect value will be EXTRA. This rule is not applicable on unanimated nouns; e.g. for قرآن plural will not be used.

Semantic features shows that a noun can be temporal (آج، ہفتہ، صبح) or spatial (لاڑکانہ، کلکتہ، گھر). These features are significant when temporal or spatial nouns occur as adjuncts in a sentence (either alone or in a postposition phrase). The order of adjuncts in a sentence is such that temporal adjuncts are followed by place (spatial) adjuncts. To encode such ordering scheme semantic features in nouns have been introduced.

**Result:** We decided on Analysis above.

**Future Work:** 1) More noun types as abstract noun may be added if found significant for syntax rules.

2) The different case forms (e.g. nominative, oblique) in temporal / place nouns such as ہفتہ، کارخانہ، لاڑکانہ، کلکتہ etc. needs to be analyzed.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	6 <sup>th</sup> November, 04	0.1	Created
Aasim Ali	28-Jun-05	5.1.0.1	Relative Pronoun added
Aasim Ali	15-Jul-05	5.1.0.2	Reflexive Pronoun added to meet temporary needs

POS ID: UPOS05

Part of Speech Name: Pronoun

Part of Speech String: pro

Lexical Entry Template:

<word>: pro, ^ PRED= 'pro', ^ NUM = {SG, PL}, ^ GEND= {M, F}, ^ PRONTYPE= {PERS, INTERROG, RELATIVE, REFLEXIVE}, ^ PERS= {1, 2, 3}, ^ FORM = {NOM, OBL}, ^ CASE= {NOM, ACC, DAT, ERG, INST, LOC\_TAK, LOC\_MAIN, LOC\_PAR}, ^ ANIM = {+, -}, ^ RESPECT = {NORESPCT, FAMILIAR, USUAL, EXTRA}, PRONFORM = <word in roman script>.

Sample Lexical Entries:

ہم: pro, ^ PRED= 'pro', ^ NUM = PL, ^ GEND= {M,F}, ^ PRONTYPE= PERS, ^ PERS= 1, ^ FORM = NOM, ^ CASE=NOM, ^ ANIM = +, ^ RESPECT = {NORESPCT, USUAL}.

ہم: pro, ^ PRED= 'pro', ^ NUM = PL, ^ GEND= {M,F}, ^ PRONTYPE= PERS, ^ PERS= 1, ^ FORM = OBL, ^ CASE={ERG, ACC, DAT, INST, LOC\_TAK, LOC\_MAIN, LOC\_PAR}, ^ ANIM = +, ^ RESPECT = {NORESPCT, USUAL}.

ہمیں: pro, ^ PRED= 'pro', ^ NUM = PL, ^ GEND= {M,F}, ^ PRONTYPE= PERS, ^ PERS= 1, ^ FORM = NOM, ^ CASE={DAT, ACC}, ^ ANIM = +, ^ RESPECT = {NORESPCT, USUAL}.

کون: pro, ^ PRED= 'pro', ^ NUM = {SG, PL}, ^ GEND= {M, F}, ^ PRONTYPE= INT, ^ PERS= 3, ^ FORM = NOM, ^ ANIM = +, ^ RESPECT = {NORESPCT, USUAL}, PRONFORM = 'KAUN'.

جو: pro, ^ PRED= 'pro', ^ NUM = {SG, PL}, ^ GEND= {M, F}, ^ PRONTYPE= RELATIVE, ^ PERS= 3, ^ FORM = NOM, ^ CASE=NOM, ^ ANIM = POS, ^ RESPECT = {NORESPCT, USUAL}, ^ PRONFORM = 'JO'.

جس: pro, ^ PRED= 'pro', ^ NUM = SG, ^ GEND= {M, F}, ^ PRONTYPE= RELATIVE, ^ PERS= 3, ^ FORM = OBL, ^ CASE={GEN, ERG, ACC, DAT, INST, LOC\_TAK, LOC\_MAIN, LOC\_PAR}, ^ ANIM = POS, RESPECT = NORESPECT, ^ PRONFORM = 'JIS'.

اپنے آپ: pro, ^ Pred= 'pro', ^ NUM = {SG, PL}, ^ GEND= {M, F}, ^ PRONTYPE= REFLEXIVE, ^ PERS= {1, 2, 3}, ^ FORM = NOM, ^ CASE=NOM, ^ ANIM = POS, ^ RESPECT = USUAL.

**Description:** Generally (but not always) pronouns stand for (*pro* + noun) or refer to a noun, an individual or individuals or thing or things (the pronoun's antecedent) whose identity is made clear earlier in the text. Pronouns are the words which are used in place of a noun. The advantage of the pronouns is that we don't have to repeat the nouns many times and it enhances the beauty of the language by reducing the repeated nouns. Relative pronoun is the pronoun, used in the relative clause, which refers to some noun/pronoun in the main clause.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reflexive pronouns are same as those in English (yourself, myself, self) when translated to Urdu (آپ، میں).

The description of the features is following:

NUM (Number) can be:

SG (singular): *Example:* میں

PL (plural): *Example:* ہم

It cannot be null or both.

GEND (Gender) can be:

M (masculine): *Example:* میں

F (feminine): *Example:* میں

It cannot be null. In Urdu, all pronoun has both values.

FORM can be:

NOM (Nominative): *Example:* میں

OBL (Oblique): *Example:* مجھ

It cannot be null or both.

CASE can be:

NOM (Nominative): *Example:* وہ

ERG (Oblique): *Example:* اس

ACC (Accusative): *Example:* اسے

Detail of other cases is present in Case Marker Document  
It cannot be null.

(UPOS01).

PRONTYPE (Pronoun Type) can be:

PERS *Example:* میں

INTERROG *Example:* کون

RELATIVE *Example:* جو

REFLEXIVE *Example:* اپنے آپ

It cannot be null or both.

ANIM (Animated) can be:

+ (Positive): *Example:* وہ

- (Negative): *Example:* وہ

It cannot be null.

RESPECT can be:

NORESPECT: *Example:* تو

FAMILIAR: *Example:* تم

USUAL: *Example:* آپ

EXTRA: *Example:* آپ (3<sup>rd</sup> Person)

It cannot be null.

PRONFORM can be: 'WOH', 'YEH', 'KIYA', 'KAUN', 'JO', 'JIS' etc.

The detail of the feature values is discussed in analysis section.

### Examples:

تو، ہم، ہمیں، مجھ، یہ، وہ، آپ، اس، کیا، کون، کہاں، کیوں، کیسا، کب، کس، کسے، جو، جس، جن، جیسے، جنہوں، جنہیں



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**POS Status:** Under Process

**Reference:** [1] مولوی عبدالحق، اردو صرف و نحو، صفحہ 41-42  
[2] Bhatt, Rajesh  
"Topics in the Syntax of the Modern Indo-Aryan Languages"  
April 4, 2003, "Correlative Clauses"  
[3] Bhatt, Rajesh  
"Topics in the Syntax of the Modern Indo-Aryan Languages"  
April 11, 2003, "Other Correlative Clauses"

**Related POS:** UPOS104, UPOS111

**Replaces:** -UPOS105

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the rule.

*Analysis1:*

Predicate of pronoun is different from other POS. Usually predicate has logical form of the word as value, but in case of pronoun only 'pro' is written. It is because Pronoun is a closed class category, and we can get a unique pronoun by only knowing its all feature values.

Number, Gender and Animated features of Pronoun have same analysis and description that is given for Noun(UPOS04).

Pronoun can be of following type.

- وہ پڑھتا ہے Personal Pronoun is used in place of a noun.

- وہ کون ہے؟ Interrogative Pronoun is also used in place of a noun for asking a question.

- وہ لڑکا آیا۔ Demonstrative Pronoun is used to indicate or specify a noun.

- جو پڑھتا ہے Personal Relative Pronoun is used in place of a noun/pronoun.

As function of Demonstrative Pronoun is different from other types of pronoun, we introduced a new POS for it (UPOS11).

Personal pronoun can have three values for person. Example of first, second and third person pronouns are میں, تم, and وہ respectively.

With reference to Form and Case, pronoun shows two type of behavior. Some pronouns do not allow a case marker or a post position. Examples of these are: اسے، جو، جسے. These pronouns have Nominative form. The other type can have a case marker/postposition after them. Examples of these are: جس، تم، تجھ، میں، اس. These pronouns have Oblique form. For description of case see UPOS01.

Form shows presence/absence of case marker/postposition after a pronoun. But different pronouns allow different case markers after them. For example, میں can come with ERG case and مجھ can come with INST case. Pronoun مجھ is used in place of مجہ therefore it does not require any case marker, but it has ACC/DAT case in it. These properties are represented by CASE feature of Pronoun.

Different pronouns are used to show different levels of respect. Usually a plural pronoun is used to show respect for a singular person. For First and Third person, pronouns has two levels of respect: "No Respect" and "Usual Respect". Examples are:

- میں آیا ہوں۔ وہ آیا ہے۔

- جس کے ہاتھ میں بیگ ہے۔

In above sentences, the subjects are 1<sup>st</sup> and 3<sup>rd</sup> person singular respectively.

- ہم آئے ہیں۔ وہ آئے ہیں۔

- جن کے ہاتھ میں بیگ ہے۔

The above sentences have two interpretations; the subjects are 1<sup>st</sup>, 3<sup>rd</sup> and 3<sup>rd</sup> person plural respectively or the sentences are used 1<sup>st</sup>, 3<sup>rd</sup> and 3<sup>rd</sup> person singular with respect.

Third person has an additional pronoun آپ for Extra respect. Second person has pronoun تم that is used for familiar persons. It is different from No Respect pronoun تو and Usual respect pronoun آپ.

All the above examples are of Personal Pronoun. Another type of pronouns is Interrogative pronouns. These are placeholders of unknown noun or pronoun in an interrogative sentence. Different interrogative pronouns are used to replace different syntactical elements of sentence. For example 'kaun' comes in place of animated noun phrase at PREDLINK position. The example sentences of interrogative pronouns, with their associated syntactical element are following.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

- (unanimated NP as predlink)؟ تمہارا نام کیا ہے؟  
(unanimated NP as object) وہ کیا پڑھ رہی تھی؟  
(animated NP as predlink)؟ وہ کون ہے؟  
(NP before 'si' in demonstrative phrase) اس نے کون سی کتاب خریدی؟  
(NP, PP or NomPP as locative adjunct) وہ کہاں گیا؟  
(PP or NomPP as locative adjunct) تم کہاں رہتے ہو؟  
(subordinate clause as XCOMP) تم کیوں آئے؟  
(adjective as predlink) تم کیسے ہو؟  
(NP as temporal adjunct) تم کب آئیں؟  
(accusative KP as object) اس نے کسے بلایا؟  
(NP as object of GP) تم نے کس کے ذریعے بات کی؟

In Urdu Yes/No questions are made by placing 'kiya' before a declarative sentence.  
کیا وہ آئے گا؟

The relative pronouns جونسا، جونسے، جونسے can be represented by جو therefore these are not included explicitly.  
Reflexive pronoun is used to accommodate the following position of اپنے آپ:

- وہ اپنے [آپ] میں گم رہتا ہے  
تم اپنی [آپ] سے باہر نکلو  
ہم اپنے [آپ] کو معاف نہیں کر سکتے

It was considered, for Reflexive Pronouns, that اپنے آپ may be entered in the Lexicon as single compound word.  
This option is temporary and chosen for its being computationally workable for the time being.

#### Analysis2:

The number, gender and person of Reflexive Pronoun does not correspond to the main Noun/Pronoun. There are only three reflexive pronouns in Urdu that are used in specific meanings always. میں is used to reflect the ego/arrogance/selfness sort of aspects, while خود اپنے آپ is used to reflect almost all reflexive pronouns of English, that is, oneself, yourself, himself, herself, myself, ourselves, etc.

**Result:** We decided on Analysis1 above.

#### Future Work:

Detailed analysis of Reflexive Pronoun.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	6 <sup>th</sup> November, 04	0.1	Created

POS ID: UPOS07

Part of Speech Name: Quantifier

Part of Speech String: quantifier

Lexical Entry Template:

<word>: quant, ^ CTYPE= {MASS, COUNT}, ^ NUM = {SG, PL}, ^ GEND = {M, F}, FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, QUANTYPE = {NORMAL, INT}.

Sample Lexical Entry:

انتے : quant, ^Pred = 'انتا', ^CTYPE=COUNT, ^ NUM = PL, ^ GEND = M, ^ FORM = {NOM,OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^QUANTYPE = NORMAL.

انتی : quant, ^Pred = 'انتا', ^CTYPE=COUNT, ^ NUM = PL, ^ GEND = F, ^NDERIVED = NEG, ^FORM = {NOM, OBL} , ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^QUANTYPE = NORMAL.

انتا : quant, ^Pred = 'انتا', ^ CTYPE=MASS, ^ NUM = SG, ^ FORM = NOM, ^ GEND = M, ^ RESPECT = NORESPECT, ^QUANTYPE = NORMAL.

انتے : quant, ^Pred = 'انتا', ^CTYPE=MASS, ^ NUM = SG, ^ GEND = M, ^ FORM = OBL, ^ RESPECT = NORESPECT, ^QUANTYPE = NORMAL.

انتی : quant, ^Pred = 'انتا', ^CTYPE=MASS, ^ NUM = SG, ^ GEND = F, ^FORM = {NOM, OBL} , ^ RESPECT = NORESPECT, ^QUANTYPE = NORMAL.

کتنا : quant, ^Pred = 'کتنا', ^ CTYPE=MASS, ^ NUM = SG, ^ FORM = NOM, ^ GEND = M, ^ RESPECT = NORESPECT, ^QUANTYPE = INT.

**Description:** Like articles, **quantifiers** are words that precede and modify nouns. They tell us how many or how much [2].

The description of the features is following:

CTYPE can be:

MASS Example: انتا

COUNT Example: بعض

NUM (Number) can be:

SG (singular): Example: انتا

PL (plural): Example: انتے

It cannot be null or both



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

GEND (Gender) can be:

It cannot be null

M (masculine): Example: اتنا

F (feminine): Example: اتنی

FORM can be:

NOM (Nominative): Example: کوئی

OBL (Oblique): Example: کسی

It cannot be null.

RESPECT can be:

NORESPECT: Example: اتنا

FAMILIAR: Example: اتنے

USUAL: Example: اتنے

EXTRA: Example: اتنے

It cannot be null.

QUANTTYPE(Quantifier Type) can be:

NORMAL: Example: اتنا

INT(Interrogative) Example: کتنا

The detail of the feature values is discussed in analysis section.

**Examples:**

کچھ ، چند، تمام، کوئی ، اتنے، ہر، سب، تھوڑا، تھوڑے، اتنا، اتنی، کئی، بعض، کل  
وہ الفاظ جن میں ایک بطور لاحقہ آتا ہے جیسے پچاس ایک، دس ایک وغیرہ  
دسیوں، بیسیوں وغیرہ  
ہزار ہا، صدہا وغیرہ  
سیر بھر، چلو بھر وغیرہ [1, p.36]

**POS Status:** Under Process

**Reference:**

[1] مولوی عبدالحق، اردو صرف و نحو، صفحہ 38 [1]

[2] عصمت جاوید، نئی اردو قواعد، صفحہ 46 [2]

[3] <http://webster.commnet.edu/grammar/determiners/determiners.htm>

**Related POS:** UPOS104

**Replaces:** -UPOS07

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the POS.

*Analysis:*

The predicate of the quantifier is its logical form. The PRED of {اتنا، اتنے} is اتنا.

CTYPE can have two values: countable and mass nouns. This property specifies that the noun following the quantifier will be a countable noun or a mass noun. If 'تمام' has a CTYPE count then the noun following it would be countable noun e.g., 'تمام لوگ'. If 'تھوڑا' has a CTYPE mass then the noun following it would be mass noun e.g., 'تھوڑا پانی'.

Number of a quantifier can be singular (SG) or plural (PL). It is a mandatory feature. If any quantifier has both singular and plural values they will be written in the form of set, e.g. {SG, PL}. Same is for Gender feature.

Form can have two values NOM and OBL. If the quantifier has a form NOM (nominative) it means the noun following that quantifier will not have any case marker/postpositions following it (the noun will not follow any case marker/postposition) e.g. 'اتنا کھانا ختم ہو گیا ہے'. If the quantifier has a form OBL (Oblique) then the noun following it will have some case marker/postpositions following it e.g., 'اتنے لڑکوں نے کھانا کھایا'.

Respect can have the four values {NORESPLECT, FAMILIAR, USUAL, EXTRA} It is same as written in the pronouns POS document (UPOS05)

**Result:** We decided on the above Analysis.

**Future Work:** Currently, we analyzed only single word quantifiers. Quantifiers consisting of two or more words will be analyzed in future.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	06 <sup>th</sup> November, 04	0.1	Created

POS ID: UPOS08

Part of Speech Name: Ordinals

Part of Speech String: ordinal

Lexical Entry Template:

<word>: ordinal, ^ PRED = '<logical form of word>', ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

Sample Lexical Entries:

پہلا : ordinal, ^Pred = 'پہلا', ^ NUM = SG, ^ GEND = M, ^ FORM = NOM, ^ RESPECT = NORESPECT.

پہلے : ordinal, ^Pred = 'پہلے', ^ NUM = SG, ^ GEND = M, ^ FORM = OBL, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

پہلے : ordinal, ^Pred = 'پہلا', ^ NUM = SG, ^ GEND = M, ^ FORM = NOM, ^ RESPECT = {FAMILIAR, USUAL, EXTRA}.

پہلے : ordinal, ^Pred = 'پہلا', ^ NUM = PL, ^ GEND = M, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

پہلی : ordinal, ^Pred = 'پہلا', ^ NUM = {SG, PL}, ^ GEND = F, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

**Description:** When objects are placed in order, we use ordinal numbers to tell their position. If ten students ran a race, we would say that the student that ran the fastest was in first place, the next student was in second place, and so on. [2] In common usage, an ordinal number is an adjective which describes the numerical position of an object, e.g., first, second, third, etc. [3]

The description of the features is following:

FORM can be:

NOM (Nominative): Example: پہلا

OBL (Oblique): Example: پہلے

It cannot be null

NUM (Number) can be:

SG (singular): Example: تیسرا

PL (plural): Example: تیسرے

It cannot be null

GEND (Gender) can be:

M (masculine): Example: دوسرا

F (feminine): Example: دوسری

It cannot be null.or both



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

RESPECT can be:

NORESPECT: *Example:* پہلا

FAMILIAR: *Example:* پہلے

USUAL: *Example:* تیسرے

EXTRA: *Example:* تیسرے

It cannot be null.

**Examples:**

پہلا، دوسرا، تیسرا، چوتھا، پانچواں، چھٹا، ساتواں، آٹھواں، ایک سو دسواں، پانچ ہزار باونواں وغیرہ  
پہلے، دوسرے، تیسرے، چوتھے، تانچویں، ساتویں وغیرہ  
تینوں، چاروں، پانچوں وغیرہ [1, p.36]  
اول، دوم، سوم وغیرہ

**POS Status:** Active

**Reference:**

- [1] "Urdu Sarf au Nahav", Maulvi Abdul Haq  
[2] <http://www.aaamath.com/nam15-ordinals.html>  
[3] <http://mathworld.wolfram.com/OrdinalNumber.html>

**Related POS:** UPOS104

**Replaces:** -UPOS08

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the rule.

*Analysis:*

The predicate of the ordinal is its logical form. The PRED of {پہلا، پہلے} is پہلا.

Number of an ordinal can be singular (SG) or plural (PL). It is a mandatory feature. If any ordinal has both singular and plural values they will be written in the form of set, e.g. {SG, PL}.

Gender of the ordinal can be masculine (M) or feminine (F). Example of masculine gender is 'پہلا' and example of feminine gender is 'پہلی'

Form can have two values NOM and OBL. If the ordinal has a form NOM (nominative) it means the noun following that ordinal will not have any case marker/postpositions following it (the noun will not follow any case marker/postposition) e.g. 'پہلا لڑکا جا چکا ہے'. If the ordinal has a form OBL (Oblique) then the noun following it will have some case marker/postpositions following it e.g., 'پہلے دو لڑکوں نے کھانا کھایا'.

Respect can have the four values NORESPECT, FAMILIAR, USUAL, EXTRA}. The respect is same as written in the pronoun document (UPOS05)

**Result:** We decided on above Analysis.

**Future Work:** We'll add more to this POS later on.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	06 <sup>th</sup> November, 04	0.1	Created

POS ID: UPOS09

Part of Speech Name: Cardinals

Part of Speech String: cardinal

Lexical Entry Template:

<word>: n, ^ PRED = '<logical form of word>', ^ NUM = {SG, PL}.

Sample Lexical Entry:

ایک: n, ^ PRED = 'ایک', ^ NUM = SG.

تین: n, ^ PRED = 'تین', ^ NUM = PL.

Description:

NUM (Number) can be:

SG (singular): Example: اتنا

PL (plural): Example: اتنے

It cannot be null or both

Examples:

ایک، دو، تین، چار بیالیس، انسٹھ، ننانوے، ہزار، دو ہزار، آٹھ سو، نو سو وغیرہ [1, p. 36]

POS Status: Active

Reference:

[1] "Urdu Sarf au Nahav", Maulvi Abdul Haq

Related POS: UPOS104

Replaces: -UPOS09

Reason: -

Replaced by: -

Reason: -



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**Analysis:** Following is the in-depth analysis of the rule.

*Analysis:*

The PRED (predicate) of the cardinal is its logical form. In case of cardinal, the PRED of cardinal is same as physical form of a word e.g., in the sample lexical entry, the PRED of 'ایک' is 'ایک' and similarly the PRED of 'تین' is 'تین'

Number of a cardinal can be singular (SG) or plural (PL). It is a mandatory feature.

**Result:** We decided on above Analysis.

**Future Work:** We'll add more to this POS later on.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	06 <sup>th</sup> November, 04	0.1	Created

POS ID: UPOS10

Part of Speech Name: Adjectives

Part of Speech String: adj

Lexical Entry Template:

<word>: adj, ^ PRED = '<logical form of word>', ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ NDERIVED = {+,-}, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

Sample Lexical Entry:

اچھا : adj, ^Pred = 'اچھا', ^ NUM = SG, ^ GEND = M, NDERIVED = NEG, ^ FORM = NOM, ^ RESPECT = NORESPECT.

اچھے : adj, ^Pred = 'اچھے', ^ NUM = SG, ^ GEND = M, ^ NDERIVED = NEG, ^ FORM = OBL, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

اچھے : adj, ^Pred = 'اچھے', ^ NUM = SG, ^ GEND = M, ^ NDERIVED = NEG, ^ FORM = NOM, ^ RESPECT = {FAMILIAR, USUAL, EXTRA}.

اچھے : adj, ^Pred = 'اچھے', ^ NUM = PL, ^ GEND = M, ^ NDERIVED = NEG, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

اچھی : adj, ^Pred = 'اچھی', ^ NUM = {SG, PL}, ^ GEND = F, ^ NDERIVED = NEG, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

ایماندار : adj, ^Pred = 'ایماندار', ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ NDERIVED = NEG, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

**Description:** Adjectives are words that describe or modify another person or thing in the sentence [2]

The description of the features is following:

NUM (Number) can be:

SG (singular): Example: اچھا

PL (plural): Example: اچھے

It cannot be null

GEND (Gender) can be:

M (masculine): Example: بڑا

F (feminine): Example: بڑی

It cannot be null



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

NDERIVED can be:

-  
It cannot be null or both

+ Example: پاکستانی  
Example: ایماندار

FORM can be:

NOM (Nominative): Example: بڑا  
OBL (Oblique): Example: بڑے  
It cannot be null

RESPECT can be:

NORESPECT: Example: اچھا  
FAMILIAR: Example: اچھے  
USUAL: Example: اچھے  
EXTRA: Example: اچھے  
It cannot be null.

**Examples:**

[1, p. 33] بیکار، جاہل، شریر، ہلکا، ٹھوس، سبز، ناشکرا، بے فکرا، سمجہ دار، بے چین، بے ڈھب، دانا، احمق، بینا، شریف، نفیس، [1, p. 33]  
خوب وغیرہ

**POS Status:** Active

**Reference:**

- [1] "Urdu Sarf au Nahav", Maulvi Abdul Haq  
[2] <http://webster.commnet.edu/grammar/adjectives.htm>

**Related POS:** UPOS104

**Replaces:** -UPOS10

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the POS.

*Analysis:*

The predicate of the adjective is its logical form. The PRED of {اچھا، اچھے} is اچھا.

NUM (number) of an adjective can be singular (SG) or plural (PL). It is a mandatory feature. If any adjective has both singular and plural values they will be written in the form of set, e.g. {SG, PL}.

Gender of an adjective can be masculine (M) or feminine (F). It can also have both values (masculine and feminine) and they are represented as {M, F}. The example of adjective having both values is 'بہادر'

NDERIVED can have two values + (positive) or – (negative). The positive value indicates that this adjective is derived from some noun e.g. NDERIVED of 'پاکستانی' is positive as the word 'پاکستانی' is derived from the word 'پاکستان', however, the NDERIVED of 'اچھا، ایماندار' is negative as it is not derived from some noun

Form can have two values NOM and OBL. If the adjective has a form NOM (nominative) it means the noun following that adjective will not have any case marker/postpositions following it (the noun will not follow any case marker/postposition) e.g. 'اچھا لڑکا چلا گیا'. If the adjective has a form OBL (Oblique) then the noun following it will have some case marker/postpositions following it e.g., 'اچھے لڑکوں نے کھانا کھایا'

Respect can have the four values {NORESPECT, FAMILIAR, USUAL, EXTRA} It is same as written in the pronouns document (UPOS05)

**Result:** We decided on the above Analysis.

**Future Work:** We'll add more to this POS later on.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	6 <sup>th</sup> November, 04	0.1	Created
Sara Hussain	17 Dec 2004	4.1.0.0	modified for compatibility with grammar file (added RESPECT feature)

POS ID: UPOS11

Part of Speech Name: Demonstrative pronouns

Part of Speech String: dem

Lexical Entry Template:

<word>: dem, ^ PRED= 'logical form of word', ^FORM = {NOM, OBL}, ^ PLACE = {NEAR, FAR}, ^ GEND = {M, F}, ^ NUM = {SG, PL}, DEMTYPE = {NORMAL, INT}, RESPECT = {NORESPECT,USUAL,FAMILIAR,EXTRA}.

Sample Lexical Entry:

وہ : dem, ^ PRED= 'وہ', ^FORM = NOM, ^ PLACE = FAR, ^GEND = {M,F}, ^ NUM = {SG,PL}, ^DEMTYPE = NORMAL.

اس : dem, ^ PRED= 'وہ', ^FORM = OBL, ^ PLACE = FAR, ^GEND = {M,F}, ^ NUM = SG}, ^DEMTYPE = NORMAL.

ان : dem, ^ PRED= 'وہ', ^FORM = OBL, ^ PLACE = FAR, ^GEND = {M,F}, ^ NUM = PL. }, ^DEMTYPE = NORMAL

یہ : dem, ^ PRED= 'یہ', ^FORM = NOM, ^ PLACE = NEAR, ^GEND = {M,F}, ^ NUM = {SG,PL}}, ^DEMTYPE = NORMAL.

ایسا : dem, ^ PRED= 'ایسا', ^FORM = NOM, ^GEND = M, ^ NUM = SG, RESPECT = NORESPECT}, ^DEMTYPE = NORMAL.

ایسے : dem, ^ PRED= 'ایسا', ^FORM = OBL, ^GEND = M, ^ NUM = SG, RESPECT = {NORESPECT,USUAL,FAMILIAR,EXTRA}}, ^DEMTYPE = NORMAL.

کیسا : dem, ^ PRED= 'کیسا', ^FORM = NOM, ^GEND = M, ^ NUM = SG, RESPECT = NORESPECT}, ^DEMTYPE = INT.

**Description:** They replace a specific noun, which has been mentioned or is obvious from context, in order to avoid repeating it. They agree in gender and number with the nouns they replace. [1]

The description of the features is following:

FORM can be:

NOM (Nominative): Example: کوئی

OBL (Oblique): Example: کسی

It cannot be null.

PLACE can be: NEAR

Example: یہ

FAR Example: وہ



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

It cannot be null or both

GEND (Gender) can be:

M (masculine): *Example:* ایسا

F (feminine): *Example:* ایسی

It cannot be null

NUM (Number) can be:

SG (singular): *Example:* ایسا

PL (plural): *Example:* ایسے

It cannot be null

DEMTYPE(Demonstrative Type) can be:

NORMAL *Example:* ایسا

INT(Interrogative) *Example:* کیسا

RESPECT can be:

NORESPECT: *Example:* ایسا

FAMILIAR: *Example:* ایسے

USUAL: *Example:* ایسے

EXTRA: *Example:* ایسے

This can be null.

The detail of the feature values is discussed in analysis section.

**Examples:**

وہ ، یہ ، اُس ، اُن ، اِس ، اِن ، ایسا ، جیسا ، کیسا وغیرہ

**POS Status:** Under Process

**Reference:** [1] مولوی عبدالحق، اردو صرف و نحو، صفحہ 47 [2] عصمت جاوید، نئی اردو قواعد، صفحہ 60

**Related POS:** UPOS104, UPOS105

**Replaces:** -UPOS11

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the POS.

*Analysis:*

The predicate of demonstrative pronoun is its logical form. The predicate of {وہ، اُس، اُن} is وہ.

In Urdu, demonstrative pronoun can be of two forms. One is Nominative (NOM) i.e. yeh یہ and other is Oblique (OBL) i.e. us اس. Nominative form is used when demonstrative pronoun is in Nominative case. When demonstrative pronoun is in some other case then the Oblique form is used along with case marker. For further details see POS Document of case markers and Grammar rule document for Case Phrase (KP).

Place of demonstrative pronouns has two values. It can be NEAR or FAR. The PLACE of 'یہ' is NEAR and the PLACE of 'وہ' is FAR.

Gender of a demonstrative pronoun can be masculine (M) or feminine (F) or it can have both the values.

Number of a demonstrative pronoun can be singular (SG) or plural (PL). It is a mandatory feature. If any demonstrative pronoun has both singular and plural values they will be written in the form of set, e.g. {SG, PL}.

**Result:** We decided on Analysis.

**Future Work:**



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	6 <sup>th</sup> November, 04	0.1	Created
Sara Hussain	18 Dec, 2004	4.1.0.0	modified for compatibility with grammar/Lexicon file (added NEG and ADVFORM features)
Sara Hussain	6 <sup>th</sup> Jan, 2005	4.1.0.1	Added SEM_TYPE feature with PLACE, TEMPORAL options

POS ID: UPOS12

Part of Speech Name: Adverbs

Part of Speech String: adv

Lexical Entry Template:

<word>: adv, ^ PRED = '<logical form of word>', ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ ADVTYPE = {VERB, ADJECTIVE, ADVERB, SENTENCE}, ^ SEM\_TYPE = {PLACE, TEMPORAL}, ^ NEG = {NEG, POS}, ^ ADVFORM = '<logical form of word in roman >'.

Sample Lexical Entry:

بہت : adv, ^Pred = 'بہت', ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ ADVTYPE = {ADJECTIVE, ADVERB}.

تھوڑا : adv, ^Pred = 'تھوڑا', ^ NUM = SG, ^ GEND = M, ^ FORM = NOM, ^ RESPECT = NORESPECT, ^ ADVTYPE = {ADJECTIVE, ADVERB}.

تھوڑے : adv, ^Pred = 'تھوڑے', ^ NUM = SG, ^ GEND = M, ^ FORM = OBL, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ ADVTYPE = {ADJECTIVE, ADVERB}.

تھوڑے : adv, ^Pred = 'تھوڑے', ^ NUM = SG, ^ GEND = M, ^ FORM = NOM, ^ RESPECT = {FAMILIAR, USUAL, EXTRA}, ^ ADVTYPE = {ADJECTIVE, ADVERB}.

تھوڑے : adv, ^Pred = 'تھوڑے', ^ NUM = PL, ^ GEND = M, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ ADVTYPE = {ADJECTIVE, ADVERB}.

تھوڑی : adv, ^Pred = 'تھوڑی', ^ NUM = {SG, PL}, ^ GEND = F, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ ADVTYPE = {ADJECTIVE, ADVERB}.

نہیں : adv, ^ NEG= POS, ^ ADVFORM = 'nahin'.

یہاں : adv, ^Pred = 'یہاں', ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ SEM\_TYPE=PLACE, ^ ADVTYPE = {SENTENCE}.

ہمیشہ : adv, ^Pred = 'ہمیشہ', ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ FORM = {NOM, OBL}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ SEM\_TYPE=TEMPORAL, ^ ADVTYPE = {SENTENCE}.

**Description** Adverbs often tell when, where, why, or under what conditions something happens or happened. Adverbs can modify verbs, adjectives, sentences and other adverbs.

The description of the features is following:



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

NUM (Number) can be:

SG (singular): *Example:* تھوڑا  
PL (plural): *Example:* تھوڑے  
It cannot be null

GEND (Gender) can be:

It cannot be null

M (masculine): *Example:* تھوڑا  
F (feminine): *Example:* تھوڑی

FORM can be:

NOM (Nominative): *Example:* تھوڑا  
OBL (Oblique): *Example:* تھوڑے  
It cannot be null

RESPECT can be:

NORESPECT: *Example:* تھوڑا  
FAMILIAR: *Example:* تھوڑے  
USUAL: *Example:* تھوڑے  
EXTRA: *Example:* تھوڑے  
It cannot be null.

ADVTYPE can be:

VERB: *Example:* آپستہ  
ADJECTIVE: *Example:* بہت  
ADVERB: *Example:* نہایت  
SENTENCE: *Example:*  
It cannot be null.

SEM\_TYPE (Semantic type) can be: PLACE:

*Example:* یہاں  
TEMPORAL: *Example:* ہمیشہ  
It can be null.

NEG can be:

NEG: *Example:* بہت  
POS: *Example:* نہیں  
This is an optional feature.

**Examples:**

آپستہ، بہت، نہایت، بہت زیادہ، کم، یکایک (خطرناک طوفان آیا)، تقریباً، زیادہ، بالکل (صحیح بات)، اکثر (وہ اکثر صحیح بات کرتے ہیں)،  
([2] اتنا (جھوٹا آدمی)، کتنا (جھوٹا شخص ہے))  
(Place) یہاں، وہاں، ادھر، ادھر (Place)  
(Temporal) ہمیشہ (Temporal)



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**POS Status:** Under Process

**Reference:**

**Related POS:** UPOS104, UPOS110

**Replaces:** -UPOS12

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the POS.

*Analysis:*

The predicate of the adverbs is its logical form. The PRED of {تھوڑا، تھوڑے} is تھوڑا.

NUM (number) of an adverb can be singular (SG) or plural (PL). It is a mandatory feature. If any adverb has both singular and plural values they will be written in the form of set, e.g. {SG, PL}.

Gender of an adverb can be masculine (M) or feminine (F). It can also have both values (masculine and feminine) and they are represented as {M, F}. The example of adverb having both values is 'نہایت'

Form can have two values NOM and OBL. If the adverb has a form NOM (nominative) it means the noun following that adverb will not have any case marker/postpositions following it (the noun will not follow any case marker/postposition) e.g. 'لڑکی تھوڑا پانی پیتی ہے'. If the adjective has a form OBL (Oblique) then the noun following it will have some case marker/postpositions following it e.g., 'تھوڑے لڑکوں نے کھانا کھایا'.

Respect can have the four values {NORESPPECT, FAMILIAR, USUAL, EXTRA} It is same as written in the pronouns document (UPOS05)

Adverb is a generic part of speech that is used in many contexts. Usually it is said to be garbage box of the POS. A word, that can not classified in any other POS, can be classified as adverb. Usually Adverbs are used to show manner of the verb, modifying an adjective, modifying another adverb. The examples are:

(Adverb modifying Verb) آہستہ چلا۔

(Adverb modifying Adjective) بہت اچھی لڑکی آئی۔

(Adverb modifying Adverb) وہ نہایت آہستہ چلا۔

Some adverbs can occur as specifiers in genitive phrases. Consider the following examples:

یہاں کے کھیت

ادھر کی آبادی

اوپر کے مکانات

ہمیشہ کی طرح

ادھر کی آبادی

Due to such nominal behavior of adverbs additional semantic feature with options of PLACE and TEMPORAL can be added. This behavior is also seen in some words (such as آگے، نیچے، اوپر) which act as nominal post positions in text also. These words can have two separate entries one in which they act as nominal post positions and the other in which they are categorized as adverbs with SEM\_TYPE of PLACE feature.

This kind of analysis will also work in postpositional phrases where the postpositions can follow adverbs or adverbial phrases. Consider the following examples:

یہاں سے

ادھر تک

اوپر میں

ہمیشہ سے

ہمیشہ کے لیے

This behavior can also be modeled by indicating PLACE / TEMPORAL in SEM\_TYPE feature in such adverbs.

**Result:** We decided on Analysis.



**Center for Research in Urdu Language Processing**  
**National University of Computer and Emerging Sciences, Lahore Pakistan**

**Future Work:**



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	06 <sup>th</sup> November, 04	0.1	Created
Sara Hussain	20 <sup>th</sup> Dec, 2004	4.1.0.0	Modified (added TENSE feature)
Sara Hussain	7 <sup>th</sup> Feb, 2005	4.1.1.2	Added FORM and _INFL_AGREEMENT feature for infinitival entry
Sara Hussain	21 <sup>st</sup> Feb, 2005	4.1.1.3	Added DEVERBAL feature for deverbals.

POS ID: UPOS13

Part of Speech Name: Verb

Part of Speech String: v

Lexical Entry Template:

<word>: v, ^ PRED = '<logical form of word>', ^ \_VERB\_FORM = '<logical form of word in roman>', ^ \_SUBJ\_CASE = {NOM, ERG}, ^ \_MORPH FORM = {root, subjunctive, perfective, habitual, infinitive, percative}, ^ TNS\_ASP HABITUAL = {POS}, ^ CAUSATION = {NO, SINGLE, DOUBLE}, ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ PERSON = {1, 2, 3}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ TENSE = {PRES, FUTURE, PAST}, ^ \_INFL\_AGREEMENT = {POSITIVE, NEGATIVE}, ^ FORM = {NOM, OBL}, ^ DEVERBAL = {ADJECTIVE, NOUN}.

Sample Lexical Entry:

ابالتا : v, ^ Pred = 'ابال<SUBJ,OBJ>', ^ \_VERB\_FORM = 'ubal', ^ \_SUBJ\_CASE=NOM, ^ \_MORPH FORM = habitual, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = NO, ^ NUM = SG, ^ GEND = M, ^ PERSON = {1, 2, 3}, ^ RESPECT = NORESPECT.

ابالتے : v, ^ Pred = 'ابال<SUBJ,OBJ>', ^ \_VERB\_FORM = 'ubal', ^ \_SUBJ\_CASE=NOM, ^ \_MORPH FORM = habitual, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = NO, ^ NUM = PL, ^ GEND = M, ^ PERSON = {1, 2, 3}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}.

ابالتے : v, ^ Pred = 'ابال<SUBJ,OBJ>', ^ \_VERB\_FORM = 'ubal', ^ \_SUBJ\_CASE=NOM, ^ \_MORPH FORM = habitual, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = NO, ^ NUM = SG, ^ GEND = M, ^ PERSON = {1, 2, 3}, ^ RESPECT = {FAMILIAR, USUAL, EXTRA}.

ابالتی : v, ^ Pred = 'ابال<SUBJ,OBJ>', ^ \_VERB\_FORM = 'ubal', ^ \_SUBJ\_CASE=NOM, ^ \_MORPH FORM = habitual, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = NO, ^ NUM = SG, ^ GEND = F, ^ PERSON = {1, 2, 3}, ^ RESPECT = NORESPECT.

ہے : v, ^ \_VERB\_FORM='hay\_pred', ^ \_MORPH\_FORM = SUBJUNCTIVE, ^ CAUSATION = NO, ^ NUM = SG, ^ GEND = {M, F}, ^ PERS = {2, 3}, ^ RESPECT = NORESPECT, ^ \_ALLOWED\_FORM = {PERFECTIVE, HABITUAL}, ^ TENSE = PRES.

کہانا : v, ^ Pred = 'کہا<SUBJ,OBJ>', ^ \_VERB\_FORM = 'kha', ^ \_SUBJ\_CASE=NOM, ^ \_MORPH\_FORM = INFINITIVE, ^ CAUSATION = NO, ^ NUM = SG, ^ GEND = M, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ FORM = NOM, ^ \_INFL\_AGREEMENT = POSITIVE.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

کھانے : v, ^Pred = 'کھا<SUBJ,OBJ>', ^\_VERB\_FORM = 'kha', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = INFINITIVE, ^CAUSATION = NO, ^ NUM = SG, ^ GEND =M, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = OBL, ^\_INFL\_AGREEMENT = POSITIVE.

کھانے : v, ^Pred = 'کھا<SUBJ,OBJ>', ^\_VERB\_FORM = 'kha', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = INFINITIVE, ^CAUSATION = NO, ^ NUM = PL, ^ GEND =M, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = {NOM,OBL}, ^\_INFL\_AGREEMENT = POSITIVE.

کھانی : v, ^Pred = 'کھا<SUBJ,OBJ>', ^\_VERB\_FORM = 'kha', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = INFINITIVE, ^CAUSATION = NO, ^ NUM = {SG,PL}, ^ GEND =F, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = {NOM, OBL}, ^\_INFL\_AGREEMENT = POSITIVE.

//entries with gender feature absent

کھانا : v, ^Pred = 'کھا<SUBJ,OBJ>', ^\_VERB\_FORM = 'kha', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = INFINITIVE, ^CAUSATION = NO, ^ NUM = SG, ^PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = NOM, ^\_INFL\_AGREEMENT = NEGATIVE.

کھانے : v, ^Pred = 'کھا<SUBJ,OBJ>', ^\_VERB\_FORM = 'kha', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = INFINITIVE, ^CAUSATION = NO, ^ NUM = SG, ^PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = OBL, ^\_INFL\_AGREEMENT = NEGATIVE.

//deverbal entry

بننا : v, ^Pred = 'بن<SUBJ>', ^\_VERB\_FORM = 'ban', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = HABITUAL, ^CAUSATION = NO, ^ NUM = SG, ^ GEND =M, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT, FAMILIAR,USUAL,EXTRA}, ^ FORM = NOM, ^DEVERBAL = {ADJECTIVE, NOUN}.

بننے : v, ^Pred = 'بن<SUBJ>', ^\_VERB\_FORM = 'ban', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = HABITUAL, ^CAUSATION = NO, ^ NUM = SG, ^ GEND =M, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = OBL, ^DEVERBAL = {ADJECTIVE, NOUN}.

بننے : v, ^Pred = 'بن<SUBJ>', ^\_VERB\_FORM = 'ban', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = HABITUAL, ^CAUSATION = NO, ^ NUM = PL, ^ GEND =M, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = {NOM,OBL}, ^DEVERBAL = {ADJECTIVE, NOUN}.

بنتی : v, ^Pred = 'بن<SUBJ>', ^\_VERB\_FORM = 'ban', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = HABITUAL, ^CAUSATION = NO, ^ NUM = {SG,PL}, ^ GEND = F, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = {NOM,OBL}, ^DEVERBAL = ADJECTIVE.

بنتی : v, ^Pred = 'بن<SUBJ>', ^\_VERB\_FORM = 'ban', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = HABITUAL, ^CAUSATION = NO, ^ NUM = SG, ^ GEND =F, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = {NOM, OBL}, ^DEVERBAL = NOUN.

بنتی : v, ^Pred = 'بن<SUBJ>', ^\_VERB\_FORM = 'ban', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = HABITUAL, ^CAUSATION = NO, ^ NUM = PL, ^ GEND =F, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = NOM, ^DEVERBAL = NOUN.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

بنتے: v, ^Pred = 'بن<SUBJ>', ^\_VERB\_FORM = 'ban', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = HABITUAL,  
^CAUSATION = NO, ^ NUM = PL, ^ GEND =F, ^PERS = {1,2,3}, ^ RESPECT =  
{NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = OBL, ^DEVERBAL = NOUN.

//the hoa word

بوا: v, ^Pred = 'بو', ^\_VERB\_FORM = 'ho', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = PERFECTIVE, ^CAUSATION =  
NO, ^ NUM = SG, ^ GEND =M, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^  
FORM = NOM, ^DEVERBAL = {ADJECTIVE, NOUN}.

بوئے: v, ^Pred = 'بو', ^\_VERB\_FORM = 'ho', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = PERFECTIVE,  
^CAUSATION = NO, ^ NUM = SG, ^GEND =M, ^ PERS = {1,2,3}, ^ RESPECT =  
{NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = OBL, ^DEVERBAL = {ADJECTIVE, NOUN}.

بوئی: v, ^Pred = 'بو', ^\_VERB\_FORM = 'ho', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = PERFECTIVE,  
^CAUSATION = NO, ^ NUM = {SG,PL}, ^GEND =F, ^ PERS = {1,2,3}, ^ RESPECT =  
{NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = {NOM,OBL}, ^DEVERBAL = ADJECTIVE.

بوئے: v, ^Pred = 'بو', ^\_VERB\_FORM = 'ho', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = PERFECTIVE,  
^CAUSATION = NO, ^ NUM = PL, ^GEND =M, ^ PERS = {1,2,3}, ^ RESPECT =  
{NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = {NOM,OBL}, ^DEVERBAL = ADJECTIVE.

بوئی: v, ^Pred = 'بو', ^\_VERB\_FORM = 'ho', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = PERFECTIVE,  
^CAUSATION = NO, ^ NUM = SG, ^GEND =F, ^ PERS = {1,2,3}, ^ RESPECT =  
{NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = {NOM, OBL}, ^DEVERBAL = NOUN.

بوئے: v, ^Pred = 'بو', ^\_VERB\_FORM = 'ho', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = PERFECTIVE,  
^CAUSATION = NO, ^ NUM = PL, ^GEND =M, ^ PERS = {1,2,3}, ^ RESPECT =  
{NORESPECT,FAMILIAR,USUAL,EXTRA}, ^ FORM = NOM, ^DEVERBAL = NOUN.

بوؤں: v, ^Pred = 'بو', ^\_VERB\_FORM = 'ho', ^\_SUBJ\_CASE=NOM, ^\_MORPH\_FORM = PERFECTIVE, ^CAUSATION  
= NO, ^ NUM = PL, ^GEND ={F,M}, ^ PERS = {1,2,3}, ^ RESPECT = {NORESPECT,FAMILIAR,USUAL,EXTRA}, ^  
FORM = OBL, ^DEVERBAL = NOUN.

## Description

Verbs (or main verbs) are used to represent actions performed.

The description of the features is following:

\_SUBJ\_CASE can be:

NOM Example: اباتنا:  
ERG Example: ابالے:

\_MORPH FORM can be:

bare: Example: اتر  
subjunctive: Example: اترے:  
perfective: Example: اتر:  
habitual: Example: اترتا:  
infinitive: Example: اترنا:  
percative: Example: اترئیے:  
It cannot be null or more than one.

TNS\_ASP HABITUAL can be:



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

CAUSATION can be:	POS: Example: اباالتنا
	NO: Example: اتر
	SINGLE: Example: اتر
	DOUBLE: Example: اتر و ا
	It cannot be null.
NUM (Number) can be:	SG (singular): Example: اتر
	PL (plural): Example: اترے
	It cannot be null or both.
GEND (Gender) can be:	M (masculine): Example: اتر
	F (feminine): Example: اتری
	It cannot be null or both.
PERSON (Person) can be:	3 (3 <sup>rd</sup> person): Example: اترے
	2 (2 <sup>nd</sup> person): Example: اتر و
	1 (1 <sup>st</sup> person): Example: اتر وں
RESPECT can be:	NORESPECT, FAMILIAR, USUAL, EXTRA
	It cannot be null. It is used for agreement.
TENSE can be:	PAST: Example: تھا
	PRES: Example: ہے
	FUTURE: Example: ہو
	This an optional feature used specifically for ' hay_dat' and ' hay_pred' conditions.
FORM can be:	NOM: Example: کھانا
	OBL: Example: کھانے
	This is an optional feature used for infinitival and deverbal entries.
_INFL_AGREEMENT (Infinitival gender agreement) can be:	POSITIVE: Example: کھانی
	NEGATIVE: Example: کھانا
	This an optional feature used specifically for infinitival verb entry. It cannot be both.
DEVERBAL can be:	NOUN: Example: بوؤں
	ADJECTIVE: Example: ہوئے
	This is an optional feature used for deverbal entries. It can be both.

The detail of the feature values is discussed in analysis section.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan



Examples:

لکھنا، کہاتا، سوئے، پڑھیں، چل

**POS Status:** Active

**Reference:**

[1] Urdu Verbs Forms and Features (MT-02\_Verb Features.doc)

[2] جامع القواعد، ابو اللیث صدیقی

[3] اردو صرف و نحو، مولوی عبدالحق

[4] نئی اردو قواعد، عصمت جاوید

**Related POS:** UPOS114, UPOS115, UPOS116

**Replaces:** -UPOS13

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the rule.

*Analysis:* About all the writers e.g. [2, 3, 4] of Urdu Grammar has categories verbs into more than one types. All of them are agreed that Verb (or Main Verb) is a type of verb that shows some action. Consider following examples:

اسے خط لکھنا پڑتا ہے۔

وہ آم کھاتا جاتا ہے۔

وہ کتاب پڑھ رہا ہوگا۔

In above examples, پڑھ, کھاتا, لکھنا are verbs (or main verbs). These verbs are used to show the action performed. The other verbs used in these sentences are used to show tense, aspect and manner of the action. For example, in second example, کھاتا is used to show the main action of sentence i.e. to eat. The other verbs جاتا and ہے are used to show continuity and present tense respectively.

Different inflectional forms of a verb can be represented using following features: Predicate, \_VERB\_FORM, \_MORPH Form, TNS\_ASP, Causation, Gender, Number, Person, and Respect. The description of these features is given in [1].

Infinitivals (verb in bare form + suffix نی, نا and نے) have been modeled as verbs having nominal properties. They introduce a feature of \_INFL\_AGREEMENT which tells whether the infinitival verb agrees with its embedded object or not (see UGR104 for details). If the agreement is NEGATIVE then the gender feature GEND is missing in that entry as can be seen in the sample entries given above. Infinitive verbs like nouns show nominative and oblique forms. For this reason feature of FORM has been introduced. This feature is also used to show different forms in deverbals.

Deverbals are verbs that can occur as adjective or noun in a sentence. DEVERBAL feature in a verb entry shows presence of deverbals as nouns or as adjective. The FORM feature gives the nominative or oblique form of a deverbal.

**Result:** We decided on Analysis.

**Future Work:**



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	06 <sup>th</sup> November, 04	0.1	Created
Sara Hussain	17 <sup>th</sup> Dec, 04	4.1.0.0	modified for compatibility with grammar/Lexicon file (changed PERSON to PERS, _MORPH FORM to _MORPH_FORM)

POS ID: UPOS14

Part of Speech Name: Aspectual Auxiliary

Part of Speech String: Asp\_Aux

Lexical Entry Template:

<word>: asp\_aux, ^ \_AUX\_FORM = '<logical form of word>', ^ JA\_TYPE = {aa, ja}, ^ \_ALLOWED\_FORM = {BARE, SUBJUNCTIVE, PERFECTIVE, HABITUAL}, ^ \_SUBJ\_CASE = {NOM, ERG}, ^ \_MORPH\_FORM = {BARE, SUBJUNCTIVE, PERFECTIVE, HABITUAL}, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = {NO, SINGLE, DOUBLE}, ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ PERS = {1, 2, 3}, ^ RESPECT = {NORESPLECT, FAMILIAR, USUAL, EXTRA}.

Sample Lexical Entry:

آنا : asp\_aux, ^ \_AUX\_FORM = 'aa', ^ JA\_TYPE = aa, ^ \_ALLOWED\_FORM = {BARE, habitual}, ^ \_SUBJ\_CASE=NOM, ^ \_MORPH\_FORM = habitual, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = NO, ^ NUM = SG, ^ GEND = M, ^ PERS = {1, 2, 3}, ^ RESPECT = NORESPECT.

آئے : asp\_aux, ^ \_AUX\_FORM = 'aa', ^ JA\_TYPE = aa, ^ \_ALLOWED\_FORM = {BARE, habitual}, ^ \_SUBJ\_CASE=NOM, ^ \_MORPH\_FORM = habitual, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = NO, ^ NUM = PL, ^ GEND = M, ^ PERS = {1, 2, 3}, ^ RESPECT = {NORESPLECT, FAMILIAR, USUAL, EXTRA}.

آئی : asp\_aux, ^ \_AUX\_FORM = 'aa', ^ JA\_TYPE = aa, ^ \_ALLOWED\_FORM = {BARE, habitual}, ^ \_SUBJ\_CASE=NOM, ^ \_MORPH\_FORM = habitual, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = NO, ^ NUM = SG, ^ GEND = M, ^ PERS = {1, 2, 3}, ^ RESPECT = {FAMILIAR, USUAL, EXTRA}.

آتی : asp\_aux, ^ \_AUX\_FORM = 'aa', ^ JA\_TYPE = aa, ^ \_ALLOWED\_FORM = {BARE, habitual}, ^ \_SUBJ\_CASE=NOM, ^ \_MORPH\_FORM = habitual, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = NO, ^ NUM = SG, ^ GEND = F, ^ PERS = {1, 2, 3}, ^ RESPECT = NORESPECT.

آئی : asp\_aux, ^ \_AUX\_FORM = 'aa', ^ JA\_TYPE = aa, ^ \_ALLOWED\_FORM = {BARE, habitual}, ^ \_SUBJ\_CASE=NOM, ^ \_MORPH\_FORM = habitual, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = NO, ^ NUM = PL, ^ GEND = F, ^ PERS = {1, 2, 3}, ^ RESPECT = {NORESPLECT, FAMILIAR, USUAL, EXTRA}.

جانا : asp\_aux, ^ \_AUX\_FORM = 'ja', ^ JA\_TYPE = ja, ^ \_ALLOWED\_FORM = {BARE, habitual, SUBJUNCTIVE}, ^ \_SUBJ\_CASE=NOM, ^ \_MORPH\_FORM = habitual, ^ TNS\_ASP HABITUAL = POS, ^ CAUSATION = NO, ^ NUM = SG, ^ GEND = M, ^ PERS = {1, 2, 3}, ^ SUBJ = NORESPECT.

Description

Aspectual Auxiliaries are used to show manner and aspect of the verb. Different inflectional forms of it can be represented using following features: \_AUX\_FORM, JA\_TYPE, \_ALLOWED FORM, \_SUBJ\_CASE, \_MORPH\_FORM, TNS\_ASP HABITUAL, CAUSATION, Gender, Number, Person and Respect. The detailed description of



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

these features is given in [1]. The feature list does not have Predicate, because it only introduces some features in Tense\_Aspct of the main verb. Aux\_Form is an alternate of Predicate that shows the name of verb. The description of the features is following:

JA\_TYPE can be:

aa Example: آنا  
ja Example: جانا  
It cannot be NULL or both

\_ALLOWED\_FORM can be:

BARE: Example: آنا  
SUBJUNCTIVE: Example: جانا  
PERFECTIVE:  
HABITUAL: Example: آتی  
It cannot be null or more than one.

\_SUBJ\_CASE can be:

NOM Example: آئے  
ERG  
It cannot be NULL

\_MORPH\_FORM can be:

BARE: Example: آ  
SUBJUNCTIVE: Example: آئے  
PERFECTIVE: Example: آیا  
HABITUAL: Example: آنا  
It cannot be NULL

TNS\_ASP HABITUAL can be:

POS: Example: آنا

CAUSATION can be:

NO: Example: آیا  
SINGLE:  
DOUBLE:  
It cannot be null.

NUM (Number) can be:

SG (singular): Example: آنا  
PL (plural): Example: آئے  
It cannot be null or both.

GEND (Gender) can be:

M (masculine): Example: آنا  
F (feminine): Example: آتی

It cannot be null or both.

PERS (Person) can be:

3 (3<sup>rd</sup> person): Example: آنا  
2 (2<sup>nd</sup> person): Example: آئے  
1 (1<sup>st</sup> person): Example: آتی

RESPECT can be:

NORESPECT, FAMILIAR, USUAL, EXTRA  
It cannot be null. It is used for agreement.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Examples:**

جاتا، آتا، پڑا، ملا، دیتا، لیتا وغیرہ

**POS Status:** Active

**Reference:**

[1] Urdu Verbs Forms and Features (MT-02\_Verb Features.doc)

[2] جامع القواعد، ابو اللیث صدیقی

[3] Miriam Butt

[4] نئی اردو قواعد، عصمت جاوید

**Related POS:** UPOS113

**Replaces:** -UPOS14

**Reason:** -

**Replaced by:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:** Following is the in-depth analysis of the rule.

*Analysis 1:* Urdu Grammar Writers have introduced a category of verbs called Helping Verbs [2]. Miriam Butt has termed these verbs as Light Verbs [3]. These verbs are used after main verb to enhance its meaning or to show some aspectual feature. These verbs are also used to make compound verbs by combining a noun or adjective and a verb. Some examples of these verbs are.

- اسے خط لکھنا پڑتا ہے۔
- وہ آم کھاتا جاتا ہے۔
- وہ کام کرتا ہے۔

In above examples, پڑتا، جاتا، کرتا are helping or light verbs. In first sentence, parta shows that the subject has to write the letter. In second sentence, jata shows the continuity of eating action. And in third sentence, karta is used after noun kaam to show action of work.

*Analysis 2:* In Analysis 1, we discussed light verbs that perform two different tasks. a) Some light verbs are used to enhance meaning of a main verb, like jata and karta used in first and second example. b) Some light verbs are used after nouns/adjective to show some action. Ismat Javed gives two different classes for these two tasks. He termed verbs of a) as Helping Verb and b) as Non-Helping Verbs [4].

This analysis seems true, because both of these categories have different purposes. Another important point is that verbs of a) can be used after verbs of b). The compound word formed by noun/adjective and a verb acts as a single noun. For example,

اس نے کہانی یاد کی۔

We can use verbs of a) after compound word kaam karna.

- اسے کہانی یاد کرنا پڑتی ہے۔
- وہ کہانی یاد کرتا جاتا ہے۔

In these examples, parta and jata are used after karma in the similar way as they are used after main verbs likhna and khana.

Another point is that we can have causation of group b) verbs but not of group a). For example, we can write

- اسے خط لکھوانا پڑتا ہے۔
- وہ کہانی یاد کروانا جاتا ہے۔

But following is not valid.

\*اسے خط لکھنا پڑواتا ہے۔

Therefore, there should be two categories, a) Aspectual Auxiliary for enhancement of verb's meaning and b) Light Verbs to use after noun/adjective in a compound word.

Different inflectional forms of an aspectual auxiliary can be represented using following features: `_AUX_FORM`, `JA_TYPE`, `_ALLOWED_FORM`, `_SUBJ_CASE`, `_MORPH_FORM`, `TNS_ASP` HABITUAL, CAUSATION, NUM (number), `GEND` (Gender), `PERS` (Person) and `RESPECT`. The description of these features is given in [1].

**Result:** We decided on Analysis 2.

**Future Work:** Some Aspectual Auxiliaries can be classified as Modals.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	6 <sup>th</sup> November, 04	0.1	Created

POS ID: UPOS15

Part of Speech Name: Light Verb

Part of Speech String: lt\_v

Lexical Entry Template:

<word>: ltv, ^ PRED = '<logical form of word>', ^\_VERB\_FORM = '<logical form of a word in roman>',  
\_SUBJ\_CASE = {NOM, ERG}, ^\_MORPH FORM = {BARE, SUBJUNCTIVE, PERFECTIVE, HABITUAL},  
TNS\_ASP HABITUAL = POS, ^CAUSATION = {NO, SINGLE, DOUBLE}, ^ NUM = {SG, PL}, ^ GEND = {M, F},  
^PERSON = {1, 2, 3}, ^ RESPECT = {NORESPLECT, FAMILIAR, USUAL, EXTRA}.

Sample Lexical Entry:

کرنا : ltv, ^Pred = 'کر<SUBJ,DPRED>', ^\_VERB\_FORM = 'kar', ^\_SUBJ\_CASE=NOM, ^\_MORPH FORM = habitual,  
^ TNS\_ASP HABITUAL = POS, ^CAUSATION = NO, ^ NUM = SG, ^ GEND = M, ^ PERSON = {1, 2, 3}, ^ SUBJ =  
NORESPLECT.

کرتے : ltv, ^Pred = 'کر<SUBJ,DPRED>', ^\_VERB\_FORM = 'kar', ^\_SUBJ\_CASE=NOM, ^\_MORPH FORM =  
habitual, ^ TNS\_ASP HABITUAL = POS, ^CAUSATION = NO, ^ NUM = PL, ^ GEND = M, ^ PERSON = {1, 2, 3}, ^  
RESPECT = {NORESPLECT, FAMILIAR, USUAL, EXTRA}.

کرتی : ltv, ^Pred = 'کر<SUBJ,DPRED>', ^\_VERB\_FORM = 'kar', ^\_SUBJ\_CASE=NOM, ^\_MORPH FORM = habitual,  
^ TNS\_ASP HABITUAL = POS, ^CAUSATION = NO, ^ NUM = SG, ^ GEND = F, ^ PERSON = {1, 2, 3}, ^  
RESPECT = NORESPECT.

کرو : ltv, ^Pred = 'کر<SUBJ,DPRED>', ^\_VERB\_FORM = 'kar', ^\_SUBJ\_CASE=NOM, ^\_MORPH FORM =  
SUBJUNCTIVE, ^CAUSATION = NO, ^ NUM = PL, ^ GEND = {M, F}, ^ PERSON = 2, ^ RESPECT = FAMILIAR.

Description

Light Verbs are used after a noun or verb to make a new verb. Different inflectional forms of it can be represented using following features: Predicate, Form, Gender, Number, Person, Respect and Causation. These features are discussed in [1].

The description of the features is following:

\_SUBJ\_CASE can be:

NOM Example: کرتا

ERG Example: کیا

It cannot be NULL

\_MORPH FORM can be:

BARE: Example: کر

SUBJUNCTIVE: Example: کرے

PERFECTIVE: Example: کیا

HABITUAL: Example: کرتا

It cannot be NULL



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

TNS\_ASP HABITUAL can be:

POS: *Example:* کرتا

CAUSATION can be:

NO: *Example:* کرتا  
It cannot be null.

NUM (Number) can be:

SG (singular): *Example:* کرتا  
PL (plural): *Example:* کرتے  
It cannot be null or both.

GEND (Gender) can be:

It cannot be null or both.

M (masculine): *Example:* کرتا  
F (feminine): *Example:* کرتی

PERSON (Person) can be:

3 (3<sup>rd</sup> person): *Example:* کرتا  
2 (2<sup>nd</sup> person): *Example:* کرتے  
1 (1<sup>st</sup> person): *Example:* کرتی

RESPECT can be:

NORESPECT, FAMILIAR, USUAL, EXTRA  
It cannot be null. It is used for agreement.

**Examples:**

کرنا، ہونا، چلنا وغیرہ

**POS Status:** Under Process

**Reference:** [1] Urdu Verbs Forms and Features (MT-02\_Verb Features.doc)  
[2] UPOS014 (Aspectual Auxiliary Document)

**Related POS:** UPOS113

**Replaces:** -UPOS15

**Reason:** -

**Replaced by:** -

**Reason:** -



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**Analysis:** Light verbs are already analyzed with Aspectual Auxiliary (MT-02\_UPOS014). The same analysis and arguments hold for Light Verbs.

Different inflectional forms of an aspectual auxiliary can be represented using following features: PRED (predicate), \_VERB\_FORM, \_SUB\_CASE, MORPH FORM, TNS\_ASP HABITUAL, CAUSATION, NUM (number), GEND (Gender), PERSON (Person) and RESPECT. The description of these features is given in [1].

**Result:** We decided on Analysis 2 of MT-02\_UPOS014

**Future Work:**



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	06 <sup>th</sup> November, 04	0.1	Created
Sara Hussain	18 <sup>th</sup> Dec, 04	4.1.0.0	modified for compatibility with grammar/ lexicon file (changed _MORPH_FORM to _MORPH_FORM, PERSON to PERS and added _ALLOWED_FORM and TENSE features)

POS ID: UPOS16

Part of Speech Name: Tense Auxiliary

Part of Speech String: tense\_aux

Lexical Entry Template:

<word>:tense\_aux, ^\_AUX\_FORM = {ho, tha, ga}, ^\_MORPH\_FORM = {SUBJUNCTIVE, PERF, BARE}, ^CAUSATION = {NO, SINGLE, DOUBLE}, ^NUM = {SG, PL}, ^GEND = {M, F}, ^PERS = {1, 2, 3}, ^RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^\_ALLOWED\_FORM = {PERFECTIVE, HABITUAL, SUBJUNCTIVE}, TENSE = {PAST, FUTURE, PRES}

Sample Lexical Entry:

تھیں : tense\_aux, ^\_AUX\_FORM = 'tha', ^\_MORPH\_FORM = PERFECTIVE, ^CAUSATION = NO, ^ NUM = PL, ^ GEND = F, ^ PERS = {1, 2, 3}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^\_ALLOWED\_FORM = {PERFECTIVE, HABITUAL}, ^ TENSE = PAST.

گا : tense\_aux, ^\_AUX\_FORM = 'ga', ^\_MORPH\_FORM = PERFECTIVE, ^CAUSATION = NO, ^ NUM = SG, ^GEND = M, ^ PERS = {1, 2, 3}, ^ RESPECT = NORESPECT ^\_ALLOWED\_FORM = SUBJUNCTIVE, ^ TENSE = FUTURE.

گے : tense\_aux, ^\_AUX\_FORM = 'ga', ^\_MORPH\_FORM = PERFECTIVE, ^CAUSATION = NO, ^ NUM = PL, ^ GEND = M, ^PERS = {1, 2, 3}, ^RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^\_ALLOWED\_FORM = SUBJUNCTIVE, ^ TENSE = FUTURE.

Description

Light Verbs are used after a noun or verb to make a new verb. Different inflectional forms of it can be represented using following features: \_Aux\_Form, Form, Gender, Number, Person and Respect. These features are discussed in [1].

The description of the features is following:

\_AUX\_FORM can be:

ho: Example: ہے

tha: Example: تھیں

ga: Example: گا

It cannot be null or more than one.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

\_MORPH\_FORM can be:

SUBJUNCTIVE: Example: ہے

PERFECTIVE: Example: تھا

BARE: Example: ہو

It cannot be null or both.

CAUSATION can be:

NO: Example: اتر

SINGLE: Example: اتار

DOUBLE: Example: اتروا

It cannot be null.

NUM (Number) can be:

SG (singular): Example: تھا

PL (plural): Example: تھے

It cannot be null or both.

GEND (Gender) can be:

M (masculine): Example: تھا

F (feminine): Example: تھی

It cannot be null or both.

PERS (Person) can be:

3 (3<sup>rd</sup> person): Example: ہو

2 (2<sup>nd</sup> person): Example: ہو

1 (1<sup>st</sup> person): Example: ہوں

It cannot be null.

RESPECT can be:

NORESPECT, FAMILIAR, USUAL, EXTRA

It cannot be null. It is used for agreement.

\_ALLOWED\_FORM can be:

SUBJUNCTIVE: Example: گا

PERFECTIVE: Example: تھا

HABITUAL: Example: تھا

It cannot be null.

TENSE can be:

PAST: Example: تھا

PRES: Example: جاتا

FUTURE: Example: ہوں

It cannot be null or more than one.

**Examples:**

ہے، ہیں، ہوں، تھا، تھی، تھے، تھیں، گا، گی، گے، ہوں، ہو



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**POS Status:** Active

**Reference:** [1] Urdu Verbs Forms and Features (MT-02\_Verb Features.doc)

**Related POS:** UPOS0113

**Replaces:** -UPOS16

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

### Analysis:

#### Analysis:

There are three undisputed tense auxiliaries: *hey*, *tha* and *ga* with their inflectional forms. They show present, past and future tense respectively. Past tense can also be shown with absence of tense auxiliary. "*ho ga*" also acts as tense auxiliary. There are two issues about it which are not agreed upon by different writers. a) Is "*ho ga*" a single word or two words coming together? It was decided that *ho* and *ga* are two separate words and *ho ga* should not be a separate word. b) Second question is what tense these words show. They can give sense of all three tenses depending on the context. As we are not dealing with context, default sense is decided to be future.

As mentioned in Verb document, tense auxiliaries are subjunctive form of *ho* or *ga* and perfective form of *tha*. Combination of two tense auxiliaries also act as tense auxiliary phrase, where first one is *ho* in bare form and second one is *ga*. This combination shows agreement on the basis of their number and person. Gender agreement is only reflected on *ga* which will be dealt in VP rule.

Different inflectional forms of a verb can be represented using following features: `_AUX_FORM`, `_MORPH Form`, Causation, Gender, Number, Person, and Respect. The description of these features is given in [1].

The feature `_ALLOWED_FORM` has been added to indicate the morpheme form of the word that occurs before this tense auxiliary.

**Result:** We decided on Analysis.

**Future Work:** Passive voice is not being dealt yet. Analysis of that may add some word in class of auxiliaries.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Omar Javed	6 <sup>th</sup> November, 04	0.1	Created
Sara Hussain	1 <sup>st</sup> March, 2005	4.1.0.1	Modified text to state that verbal nouns are special type of nouns (not adjectives)

POS ID: UPOS117

Part of Speech Name: Verbal Noun

Part of Speech String: v\_n

Lexical Entry Template:

<word>: v\_n, ^ PRED = '<logical form of word>and Subcategorization Frame', ^ ACTION\_TYPE = {ker, ho, ...},

Sample Lexical Entry:

خاموش:v\_n, ^ PRED = 'خاموش<SUBJ,OBJ>', ^ ACTION\_TYPE = ker.

خاموش:v\_n, ^ PRED = 'خاموش<SUBJ>', ^ ACTION\_TYPE = ho.

**Description**

Verbal nouns are special type of nouns that are used to make a complex predicate. These nouns has verb like properties hence they have Sub-categorization Frame.

**Examples:**

یاد، شروع، محسوس، وغیرہ



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**POS Status:** Under Process

**Reference:**

- [1] عصمت جاوید، نئی اردو قواعد، صفحہ 122
- [2] Miriam Butt, Structure of Complex Predicate in Urdu
- [3] Miriam Butt, Discussion at EGD\_ULP Meetings

**Related POS:** UPOS115

**Replaces:** -UPOS17

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:**

Verbs in Urdu (and other Indo Aryan languages) are different from verbs of major languages of world like English and Arabic. Common languages have a word for almost all actions. But in Urdu most of the actions are described by a compound of noun/adjective and a light verb. The examples are: شروع کرنا، یاد کرنا:

The nouns in above examples are special type of nouns as they have sub-categorization frame. For example, 'yaad' requires Subject and Object as in the example:

اس نے سبق یاد کیا۔

A verbal noun can have multiple sub-categorization frame depending upon the light verb used after it. For example:

اس نے ہجوم کو خاموش کیا۔

ہجوم خاموش ہوا۔

**Result:** we decided on above analysis.

**Future Work:** Currently we do not allow features to restrict a verbal noun to some type of light verbs only. For example, we can say 'yaad aayi' but not 'shroo aayi'. These semantics will be dealt in future.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Tafseer Ahmed	5th Oct, 04	4.0.0.1	Created
Sara Hussain	18 <sup>th</sup> Dec, 04	4.1.0.0	modified for compatibility with grammar file (changed Part of Speech String from intns to intens)

POS ID: UPOS118

Part of Speech Name: Intensifier

Part of Speech String: intens

Lexical Entry Template:

<word>: intens, ^INSFORM = 'بھی' , ^INSTYPE = {ADVERB,NOUN,VERB}.

Sample Lexical Entry:

بھی: intens, ^INSFORM = 'بھی' , ^INSTYPE = {NOUN, VERB} .

Description

Intensifiers are used to intensify the meaning of an adverb, noun or verb.

Examples:

بھی، بھئی

POS Status: Under Process

Reference:

Related POS:

Replaces: -

Reason: -

Replaced by: -

Reason: -



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**Analysis:**

Intensifiers are used to intensify the meaning of an adverb, noun or verb. Usually these are part of adverbial phrase, noun phrase or verbal phrase. The examples of intensifier are:

حامد ہی آئے گی۔

بہت ہی اچھی لڑکی آئی۔

**Result:** We agreed on above analysis.

**Future Work:** Words like وہی are formed by merging of وہ and ہی. The POS of وہی is to be determined.



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Tafseer Ahmed	6th Oct, 04	4.0.0.1	Created

POS ID: UPOS118

Part of Speech Name: wala

Part of Speech String: wala

Lexical Entry Template:

<word>: wala, ^ PRED= '<logical form of word>and Subcategorization Frame', ^ NUM = {SG, PL}, ^ GEND = {M, F}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ FORM = {NOM, OBL}.

Sample Lexical Entry:

والا: intns, ^PRED = 'والا<OBJ>', ^ NUM = SG, ^ GEND = M, ^ RESPECT = NORESPECT, ^ FORM = NOM.

Description

Wala is special part of speech that has only one word. Wala is use to make a specifier for noun.

Examples:

والا

POS Status: Under Process

Reference:

Related POS: UPOS104, UPOS105

Replaces: -

Reason: -

Replaced by: -

Reason: -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:**

Wala is used in phrases to specify noun. Wala is used to show many relations. For example:

نیلے والی کتاب (The book of blue color)

ٹھیلے والا آدمی (The man who owns the cart)

باغ والا درخت (The tree that is present in the garden)

Another usage of wala phrase is in copular sentences. It shows future in this context. For example:

وہ آنے والا ہے۔ (He is about to come)

Wala is a unique word because it does not seem like any other word of Urdu. It seems to be similar of genitive case marker 'ka', but no grammar writer consider it as a case marker. It is also seems similar to postpositions, as it follows a noun phrase, but do not have directional, locative or temporal semantics like postpositions. Historically, it is not included in the list of postpositions provided by grammar writers. It is why 'wala' is made a new POS.

**Result:** We agreed on above analysis.

**Future Work:**



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Tafseer Ahmed	5th Oct, 04	4.0.0.1	Created

POS ID: UPOS120

Part of Speech Name: Genitive Pronoun

Part of Speech String: gen\_pro

Lexical Entry Template:

<word> : gen\_pro, ^ PRED= '<logical form of word>and Subcategorization Frame', ^ NUM ={SG,PL}, ^ GEND = {M,F}, ^ RESPECT = {NORESPECT, FAMILIAR, USUAL, EXTRA}, ^ FORM = {NOM, OBL}, ^PERS =3.

Sample Lexical Entry:

تمہارا : gen\_pro, ^Pred = 'تمہارا', ^ NUM = SG, ^ GEND= M, ^ FORM = NOM, ^ RESPECT = NORESPECT, ^ PERS=2.

Description

Genitive Pronouns are special kind of pronoun that are used as specifier. These are present in genitive case.

Examples:

میری، ہماری، تمہارا

POS Status: Under Process

Reference: UPOS105 (Pronoun Document)

Related POS: UPOS105

Replaces: -

Reason: -

Replaced by: -

Reason: -



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**Analysis:** Following is the analysis of this topic.

*Analysis 1:* Genitive Pronouns are the pronouns that are in genitive form. As these can occur only at specifier position, and can not act as subject and object (the usual position of pronouns), these are classified as a POS. The features of genitive pronouns are similar to prnouns as discussed in UPOS105 document.

*Analysis 2:* Genitive pronouns are normal pronouns which have CASE = GENITIVE as a feature.

**Result:** We agreed on analysis 1.

**Future Work:**



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Sara Hussain	21 <sup>st</sup> Dec, 2004	4.1.0.0	Created
Aasim Ali	29-Jun-05	4.1.0.1	Removed irrelevant (sub_conj) items

POS ID: UPOS121

Part of Speech Name: Coordinate Conjunction

Part of Speech String: coord\_conj

Lexical Entry Template:

<word>: coord\_conj, CONJ\_FORM = '< logical form of word in roman>', ^ NUM = {SG,PL}.

Sample Lexical Entry:

اور:coord\_conj, CONJ\_FORM = 'AUR', ^ NUM = PL .

یا:coord\_conj, CONJ\_FORM = 'YA'.

**Description:** A coordinate conjunction is used to join phrases and/or sentences.

The description of the features is following:

NUM (Number) can be:

SG (singular): Example: لڑکی

PL (plural): Example: لڑکیاں

This is an optional feature. It cannot be both.

The detail of the feature values is discussed in analysis section.

**Examples:**

اور، یا



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**POS Status:** Active

**Reference:**

[1]

[2]

**Related POS:**

**Replaces:** - NULL

**Reason:** -

**Replaced by:** -

**Reason:** -



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**Analysis:** Following is the in-depth analysis of the rule.

*Analysis:* Number feature is used in conjunction 'AUR' to add plural sense to phrase/sentence that contains it.

**Result:** We decided on Analysis above.

**Future Work:**



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Sara Hussain	21 <sup>st</sup> Dec, 2004	4.1.0.0	Created

POS ID: UPOS121

Part of Speech Name: Comma

Part of Speech String: comma

Lexical Entry Template:

<word>: comma.

Sample Lexical Entry:

⌘: comma.

**Description:** A comma is used to separate different parts of a sentence. It is also used to separate more than two phrases where the last phrase is separated by a conjunction.

Examples:

‘

**POS Status:** Active

**Reference:**

[1]

[2]

**Related POS:**

**Replaces:** - NULL

**Reason:** -

**Replaced by:** -

**Reason:** -



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**Analysis:** Following is the in-depth analysis of the rule.

*Analysis:* NULL

**Result:**

**Future Work:**



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Sara Hussain	3 <sup>rd</sup> March, 2005	4.1.0.0	Created

POS ID: UPOS123

Part of Speech Name: Verbal modifier

Part of Speech String: v\_adx

Lexical Entry Template:

<word>: v\_adx, ^ PRED = '<logical form of word>and Subcategorization Frame', ^ ACTION\_TYPE = {ker, ho, ...},  
^ NUM = {SG, PL}, ^ GEND = {M, F}.

Sample Lexical Entry:

ہلکا : v\_adx, ^Pred ='ہلکا<SUBJ,OBJ>', ^ NUM = SG, ^ GEND = M, ^ ACTION\_TYPE = ker.

ہلکے : v\_adx, ^Pred ='ہلکا<SUBJ,OBJ>', ^ NUM = PL, ^ GEND = M, ^ ACTION\_TYPE = ker.

ہلکی : v\_adx, ^Pred ='ہلکا<SUBJ,OBJ>', ^ NUM = {SG,PL}, ^ GEND = F, ^ ACTION\_TYPE = ker.

Description

Verbal modifiers are special type of adjectives / adverbs that are used to make a complex predicate. These modifiers have verb like properties hence they have Sub-categorization Frame.

The description of other features is as follows:

NUM (Number) can be:

SG (singular): Example: ہلکا  
PL (plural): Example: ہلکے  
It cannot be null.

GEND (Gender) can be:

M (masculine): Example: ہلکا  
F (feminine): Example: ہلکی

It cannot be null or both.

Examples:

پورا، ہلکا، طلب، عطا وغیرہ



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**POS Status:** Under Process

**Reference:**

**Related POS:** UPOS115

**Replaces:** -UPOS17

**Reason:** -

**Replaced by:** -

**Reason:** -



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

**Analysis:**

Verbs in Urdu (and other Indo Aryan languages) are different from verbs of major languages of world like English and Arabic. Common languages have a word for almost all actions. But in Urdu most of the actions are described by a compound of noun/adjective/ adverb and a light verb. The examples are: شروع کرنا، یاد کرنا، شروعات کرنا  
This class is different from verbal noun (v\_n) since it shows agreement with semantic object of the sentence. However verbal nouns do not show such an agreement. The tables below show this agreement.

Itv	v_adx (adjective)	KPmain (object)	KPmain (subject)
کرنا	صاف	میز	وہ
کرنا	اونچی	میز	وہ
کرنا	اونچی	میزیں	وہ
کرنا	اونچا	ہاتھ	وہ
کرنا	اچھی	تقریر	وہ
رکھنا	ہلکا	بستہ	وہ
رکھنا	ہلکے	بستے	وہ

Itv	v_adx (quantifier)	KPmain (object)	KPmain (subject)
کرنا	پوری	خواہش	وہ

Itv	v_adx (adverb)	KPmain (object)	KPmain (subject)
کرنا	آہستہ	گاڑی	وہ
کرنا	آہستہ	گاڑیاں	وہ

**Result:** we decided on above analysis.

**Future Work:**



Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan

Reference No:

Revision History:

Name	Change Date	Version	Description of Changes
Aasim Ali	14 <sup>th</sup> June 2005		Created

POS ID: UPOS124

Part of Speech Name: Subordinate Conjunction

Part of Speech String: sub\_conj

Lexical Entry Template:

<word>: sub\_conj, CONJ\_FORM = '< logical form of word in roman>', ^ NUM = {SG,PL}.

Sample Lexical Entry:

کہ:sub\_conj, CONJ\_FORM = 'KAY'.

کیونکہ:sub\_conj, CONJ\_FORM = 'KIYONKAY'.

**Description:** A subordinate conjunction is used to join phrases and/or sentences.

The description of the features is following:

NUM (Number) can be:

SG (singular): Example: لڑکی

PL (plural): Example: لڑکیاں

This is an optional feature. It cannot be both.

The detail of the feature values is discussed in analysis section.

**Examples:**

کہ، کیونکہ



**Center for Research in Urdu Language Processing  
National University of Computer and Emerging Sciences, Lahore Pakistan**

**POS Status:** Active

**Reference:**

[1]

[2]

**Related POS:**

**Replaces:** - NULL

**Reason:** -

**Replaced by:** -

**Reason:** -



**Center for Research in Urdu Language Processing**  
**National University of Computer and Emerging Sciences, Lahore Pakistan**



**Analysis:** Following is the in-depth analysis of the rule.

*Analysis:* Yet to be done.

**Result:** We decided on Analysis above.

**Future Work:**