A Computational Classification of Urdu Dynamic Copula Verb

Qaiser Abbas
Fachbereich Sprachwissenschaft
Universität Konstanz
78457 Konstanz, Germany

Ghulam Raza
Department of Computer & Information Sciences
Pakistan Institute of Engineering & Applied Sciences
Islamabad, Pakistan

ABSTRACT

In this paper, a lexical functional grammar for an automatic classification of Urdu copula verb hO (be/become) is presented according to linguistic theories. A test suite of sentences containing almost all different conjugation forms of copula verb is extracted from a raw corpus. It is tried to keep only the cases of copular construction because the copula verb hO is very much dynamic in nature of function. The respective syntactic and functional structures of different cases of copular construction are presented, through which the lexical, syntactical and functional information required by copula verb is explored. The explorations made computationally are then compared with the existing linguistic theories as a proof of evaluation. It is an attempt to classify the state-of-the-art dynamic behavior of Urdu copula verb.

General Terms:

Copula Verbs, Subject, Predicate Link, Lexical Functional Grammar

Keywords:

Computational classification, Urdu copula verb, Complex predicate, Infinitive, Non-aspectual, C & F structure

1. INTRODUCTION

Verb classification is an important issue in computational linguistics (CL) to understand the role of verbs in semantics. This concept is involved in many tasks like semantic parsing, semantic search, information extraction, etc. The functional structure depends on the verb categories and an automatic predicate argument identification is not possible without having information of verbs. Due to this, VerbNet [26] and FrameNet [4] are being used extensively in extraction of predicate argument structure.

These resources like VerbNet and FrameNet are not used to solve the related problem mentioned above, rather an own test suite for the Urdu language is built because Urdu is an under resource language. A lexical functional grammar (LFG) is coded with candidate lexical items using a XLE[1] parser environment that provides a framework to exploit the syntactical and semantical information in it. After getting predicate argument structures, these

are compared with existing linguistic theories for validation along with some novel ideas for the said copula verb hO (be/become) of Urdu language. This whole work is exercised to build a classification/categorization for the dynamic copula verb. A similar kind of work was performed for extraction of different types of modal verbs for Urdu language by Abbas et. al. in [3].

For development of a state-of-the-art computational classification of copula verb, a limited test suite of 200 sentences is collected. The test suite is completely covered the lexicon of the grammar developed. The experiments are carried out with different types of sentences having copula verb. The parsed sentences obtained after experimentation are compared with existing theories and information gathered from native speakers.

The copula verb can express the functions of copular predication, static vs dynamic expression, tense auxiliary and as a light verb [7, 6] in verb complex predicates [5]. The examples of these functional expressions are divided into different conjugation forms which are nineteen in number. These forms contained infinitives, non aspectual and aspectual [10] ones. These different conjugation forms are given in table 1. Through this computational study of copula verb, eleven different predicate argument structures are classified which are discussed along with their syntactic and functional requirement in section 3. Our experimental methodology is illustrated in section 2 and finally, the conclusion and the future work is presented in section 4.

2. EXPERIMENTAL METHODOLOGY

Relevant linguistic theories and the native knowledge of Urdu language are the two measures considered to build a classifier in the form of a LFG. These theories are discussed with proposed classified predicate argument structures in section 3. This classified LFG is coded in an IDE (integrated development environment) of XLE parser[2]. The labeling of predicates is adopted with minor changes from an annotation scheme of treebanks [1, 2] existed for Urdu. The test suite sentences are parsed on the classifier and the predicate argument structures of copula verb are identified which are discussed in section 3. The respective c (syntactic structure) and f (functional structure) structures concluded by the classifier along with their example sentences are also given in that section.

---

[1]This parser was developed by Xerox Palo Alto Research Center, USA.

[2]http://www.parc.com/

Table 1. Conjugation forms of verb *hO*

Conjugation forms	Feature description
hO-nA	be.Inf.M.Sg
hO-nE	be.Inf.M.Pl
hO-nI	be.Inf.F
tHA	be.M.3Sg.Past
tHE	be.M.3Pl.Past
tHI	be.F.3Sg.Past
tHIN	be.F.3Pl.Past
hE	be.2/3Sg.Pres
hEN	be.Pl.Pres
hO	be.2.Pres
	be.2/3Sg.Subjn
hON	be.1Sg.Pres
	be.1Sg/Pl.Subjn
hUA	be.Perf.M.Sg
hUE	be.Perf.M.Pl
hUI	be.Perf.F.Sg
hUIN	be.Perf.F.Pl
hOtA	be.Imperf.M.Sg
hOtE	be.Imperf.M.Pl
hOtI	be.Imperf.F.Sg
hOtIN	be.Imperf.F.Pl

3. CLASSIFICATION OF COPULA VERB

The design of the copula verb is divided according to their well known conjugation forms as depicted in table 1 and it is tried to embed some more argument structures to example sentences in the test suite. The use of the copula verb is divided into six different categories. Among these categories, three are the major ones e.g. infinitive, non-aspectual and aspectual. The other three includes the use of copula verb as light verb, participle and miscellaneous. The copular construction is divided according to enrichment of argument structures and the individual cases of these categories as listed as follows:

3.1 Infinitive Copula

The infinitive copula *honA* (to be/ to become) has three different conjugation forms in Urdu, which are *honA* (sg, masc), *honE* (pl, masc) and *honI* (sg, fem). In modern Urdu, the last conjugation form *honI* is also being in use for plural number of feminine gender. The two example sentences with conjugation form *honA* and *honE* are presented in examples 1 and 2 and their respective analysis of c & f structures are given in figure 1 and 2 respectively.

(1) mujHE   lambA   honA
     I.masc+fem.sg=Dat long.masc.sg be.cop.masc.sg
     hE.
     be.pres.masc+fem.sg.

     'I have to become long.'

Both the examples 1 and 2 are the case of dynamic readings. In example 1, the subject (SUB) of the sentence *mujHE* has a dative case (DAT) of personal pronoun (PPERS). The predlink (PRD) of the sentence *lambA* is an adjective (ADJ). In these conjugation forms of infinitive copula verb, an adjective phrase (ADJP) is considered only for discussion here. The other phrases like noun phrase with nominative case (NPnom), spatial or temporal postpositional phrase (PPspt or PPtmp), adverbial phrase (ADVP), etc are also possible which will be discussed in upcoming sentences in this paper. The subject NPdat has no effect on number and gender agreement except the default masculine singular form of infinitive copula

due to dative case as depicted in figure 1. In literary Urdu, example 1 can be spoken by a feminine gender and then it would be like this as *mujHE lambI honA hE*. Even the sentence in example 1 is fairly precise and normal in spoken by a dative SUB having feminine gender, but again there is no any effect of adjectival PRD's number and gender on default agreement of infinitive copula. If the PRD has an oblique case, then the agreement of infinitive copula remain in default state.

(2) un=kE   bAl   lambE
     Their.masc+fem.pl=POSS hair.masc.pl long.masc.pl
     honE   hEN.
     be.cop.masc.pl be.pres.masc+fem.pl.

     'Their hairs have to become long.'

In example 2, the situation is different. Due to nominative case (NOM) of SUB, the agreement of SUB *bAl* applies on PRD *lambE*, VCOP *honE* and tense auxiliary (VAUX) *hEN* simultaneously as can be seen in c and f structure of figure 2. However, in case of feminine gender of SUB, two conventions of infinitive copula are being ruled in the native society. First is with *honIN* (pl,fem) *hEN* (pl,masc+fem) and the second is with *honI* (sg+pl, fem) *hEN* (pl, masc+fem). Both are considered precise and accurate. The possible rule for an infinitive copula having various options of agreement is given in appendix A. All the feature structure values are common except window, which means the lexical items can be used before and after the predicate of a sentence.

3.2 Non Aspectual Copula

Non aspectual copular construction is divided into the following categories.

3.2.1 Existential Copula. The existential copular constructions are those in which the verb *hO* would be used as main verb with existential reading. All tense form e.g. present, past and future are possible in existential constructions. In these copular constructions, only the SUB is the argument of the verb that agrees and all other things are modifiers e.g. NPmod, PPmod, ADVP, etc. The modifier and an adverb can be temporal, spatial, etc. The grammar is checked with or without modifiers and a sentence with modifier is presented here in example 3.

(3) kal   a2Id   hE.
     Tomorrow festival.fem.sg be.cop.pres.masc+fem.sg

     'It is Eid(festival) tomorrow.'

In Urdu, the present tense can be used for tomorrow's event. If some person ask a question like *What is the day tomorrow?* , then answer of this is *It is Eid(festival) tomorrow*. The future tense *hO gA* (will/shall be) can also be used but then the meaning would be changed into may/may not. Here, in this example when a present tense will be used then the meaning of confirmation is there. The respective c and f structure of example sentence is given in figure 3.

In this construction, a noun phrase NPmod is acting like a modifier MODF and its type is temporal noun as can be seen in f-structure. Being concise in discussion, a rule with different possibilities is given in appendix B. Since the existential copular construction is the case of intransitive use of verb, so the rule is very simple in its nature and there is no any existence of PRD in this construction. Due to this, It is still a hot issue that in these type of constructions, the verb hO is copula or not. However, only those sentences are selected which are the true cases of existential copular construction.

```
CS 1:        S:41
           /    |    \
     NPdat:21 ADJP:33  VCMAIN:40
        |       |        / \
     PPERS:2  ADJ:4  VCOP:6 VAUX:8
        |       |       |     |
     mujHE:1 lambA:3 honA:5  hE:7
```

"mujHE lambA honA hE"

```
   [PRED  'honA<[1:mujHE], [3:lambA]>'                                ]
   |        1[PRED 'mujHE'                                       ]    |
   |SUB   2|CASE DAT, NUM sg, PERS 1, TYPE personal-pronoun|    |
  7|     21[                                                   ]    |
  8|                                                                 |
  5|        3[PRED 'lambA'                              ]           |
  6|PRD   4|GEND masc, NUM sg, TYPE adjective|           |
 40|     33[                                              ]           |
 41[FORM infinitive, GEND masc, MOOD indicative, NUM sg, TENSE pres]
```

Fig. 1.   The structure analysis of example 1

```
CS 1:              S:64
              /      |       \
       NPnom:69   ADJP:56   VCMAIN:63
        /   \        |        /    \
   KPposs:27  N:6  ADJ:8  VCOP:10 VAUX:12
     /  \      |     |       |       |
 PPERS:2 CM:4 bAl:5 lambE:7 honE:9  hEN:11
    |     |
   un:1  kE:3
```

"un kE bAl lambE honE hEN"

```
   [PRED  'honE<[5:bAl], [7:lambE]>'                                         ]
   |        [PRED 'bAl'                                                  ]   |
   |        |          [PRED 'kE'                                   ]   |   |
   |        |          |     {1[PRED 'un'                        ]}  |   |   |
   |SUB   |SPEC { 3|NOM {2[CASE NOM, NUM pl, PERS 3, TYPE personal-pronoun]} } |   |
   |     5|        4|                                            |   |   |
   |     6|       27[CASE POSS, GEND masc                        ]   |   |
 11|     69[CASE NOM, GEND masc, NUM pl, PERS 3, TYPE count-noun ]   |   |
 12|                                                                      |
  9|        7[PRED 'lambE'                          ]                    |
 10|PRD   8|GEND masc, NUM pl, TYPE adjective|                    |
 63|     56[                                          ]                    |
 64[FORM infinitive, GEND masc, MOOD indicative, NUM pl, TENSE pres       ]
```
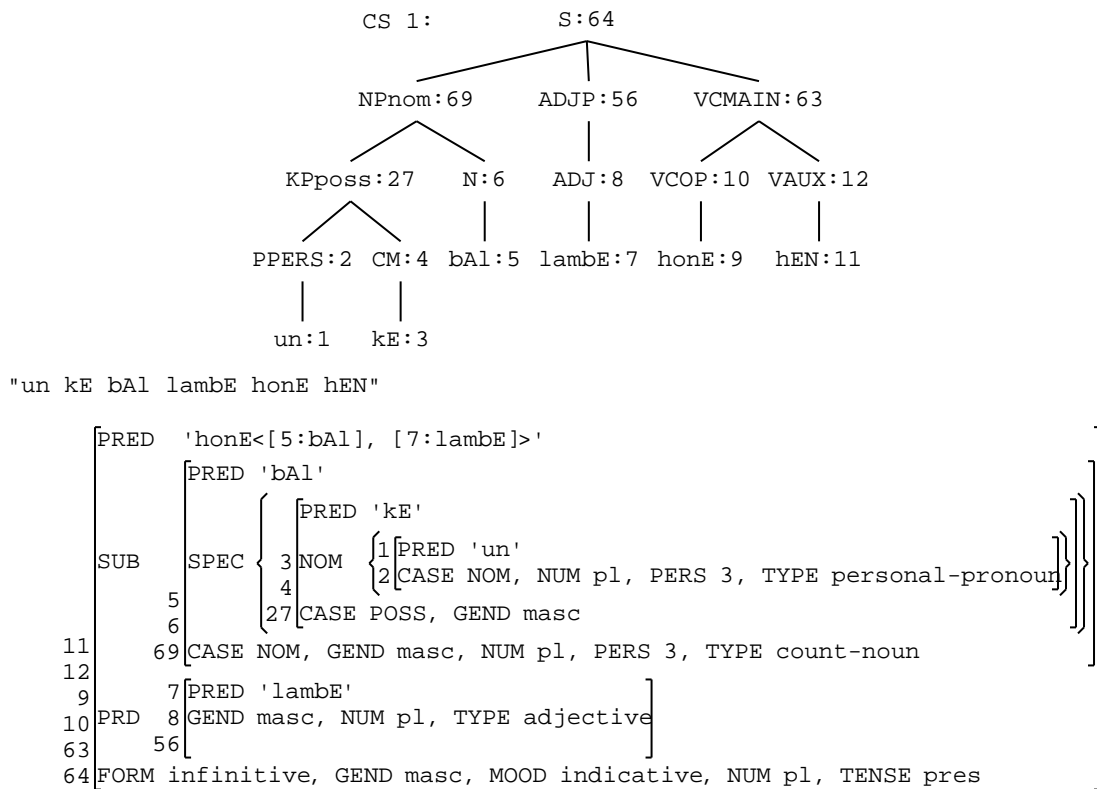
Fig. 2.   The structure analysis of example 2

3.2.2 Identification & Classification Copula. The meaning of copular construction is diverse and complex as narrated by De-clerck in [16] and Hengeveld in [19]. Similarly, Curnow in [14] divided the identification and classification meaning of copular con-

```
                                   "kal a2Id hE"

  CS 1:      S:42                        ┌PRED  'hE<[3:a2Id]>'                                      ┐
                                         │     3┌PRED 'a2Id'                                    ┐   │
                                         │SUB  4│CASE NOM, GEND fem, NUM sg, PERS 3, TYPE temporal-noun│
     NPmod:18 NPnom:36 VCMAIN:40         │    36└                                               ┘   │
                                         │      ⎧ 1┌PRED 'kal'                              ⎫        │
        │       │        │              5│MODF  ⎨ 2│GEND masc, NUM sg, PERS 3, TYPE temporal-noun⎬    │
       N:2     N:4     VCOP:7           7│      ⎩18└                                       ⎭        │
                                       40│                                                           │
        │       │        │            42└GEND fem, MOOD indicative, NUM sg, TENSE pres               ┘
      kal:1   a2Id:3    hE:5
```
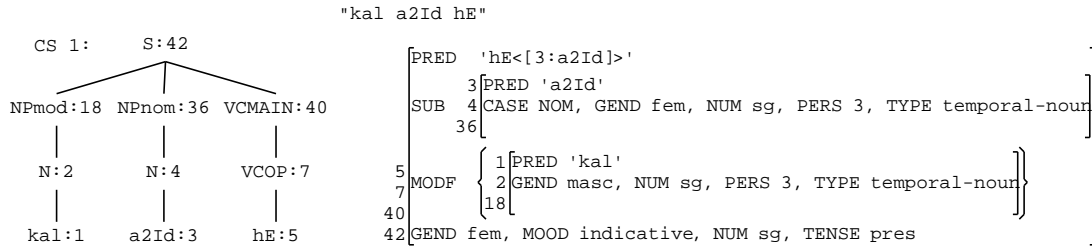
Fig. 3. The structure analysis of example 3

struction in four ways which are verbal copula, particle copula, zero copula and inflectional copula. In case of Urdu, the verbal copula is used in meaning of identification and classification as shown in example 4 and 5 respectively.

(4) woh laRkA   mErA   bHAI
     That boy.masc.sg my.masc.sg=POSS brother.masc.sg
     hE.
     be.cop.pres.masc+fem.sg

     'That boy is my brother.'

(5) ye kItAb   hE.
     This book.fem.sg be.cop.pres.fem+masc.sg

     'This is a book.'

In Raza's thesis [25], it is claimed on the basis of Declerck's theory [16] that the verbal copula in Urdu can have the meaning of specification and predication e.g. in example 4 'that boy' is specified by 'my brother' and in example 5 'the book' is predicated to 'this'. Further analysis of identification and classification is elaborated in figure 4. Only example 4 is presented because the other has the same concept. In this example, the SUB *laRkA* has no case marker, hence agrees with the copula verb *hE* in gender and number. Modifiers can exist at any position before the copula verb. These modifiers can be spatial. instrumental, temporal etc and its feature structure with almost all possibilities is very much similar to appendix A.

3.2.3 Attributive Copula. Attributive copula are those who attribute some property to subject of a sentence. These attributes are normally adjectives and it can also be a possessive. Adjectives attributes are exactly similar to example 2, if infinitive copula does not exist in that sentence. The possessive attribution gives a meaning of 'of' like in English and discussed by Raza in [24]. The examples of adjective and possessive attributes are given in example 6, 7 and 8.

(6) mErE   bAl   lambE
     My.masc.pl=POSS hair.masc.sg+pl long.masc.pl
     hEN.
     be.cop.pres.masc+fem.pl

     'My hairs are long.'

(7) mErI   laRkIyAN lambI
     My.fem.sg+pl=POSS girls.fem.pl long.fem.sg+pl
     tHIN.
     be.cop.past.fem.pl

     'My girls were long.'

(8) nIdA   sOlah sAl=kI
     Nida.fem.sg sixteen years=POSS
     hE.
     be.cop.pres.fem+masc.sg

     'Nida is of sixteen years.'

In example 6, 'long' is attributed to 'my hairs'. Similarly, in example 7, 'long' is attributed to 'my girls' but due to past tense this attribution is not confirmed in present situation that either this attributed property is existed or not. *laRkIyaN* is the SUB which agrees with the copula verb *tHIN* and the SUB has also an agreement effect on PRD if it is an ADJP. Finally, 'of sixteen years' is totally different from adjective case of attribution. Here, it is a possessive attributed property to SUB, which means 'sixteen years' are attributed to *Nida*, which is a SUB that agrees with copula verb. The PRD's possessive case marker is also in agreement with the SUB. The analysis of examples 7 and 8 are given in their respective c & f structures in figure 5 and 6 respectively.

The analysis in figure 6 is incomplete because it is giving us only one classification of predicate argument structure and not the other one. The PRD in this sentence should be XCOMP-PRD, where the PRED *kI* will take an argument *sAl* as an object OBJ and *solah* should be the specifier SPEC of PRED *sAl*. This is the correct analysis which will be performed in future. However, it is dropped here because it is the case of possessive copula construction.

The modifiers can appear at any position in adjectival attribution with no any effect at argument level but if a MODF appears in a possessive attribution, then it can become an argument of the sentence. For example, a MODF of nominal NOM type after PRD can make the whole PRD as of NOM type e.g. *solah sAl kI bacI* (a kid of sixteen years), an instrumental modifier can change the whole story e.g. *solah sAl ki laRki Se* (from a girl of sixteen years) and similarly *a2Umr Se solah sAl ki* (of sixteen years by age) and so on. So, we can say that this possessive attribution construction is restrictive as compared to examples 6 and 7. The rule of possibilities for adjectival attribution is similar to appendix A and the rule for possessive attribution can be seen in appendix C.

3.2.4 Spatial Copula. As described by Butt et. al. in [8], the location provides a spatial position of an entity and hence it is PRD. Due to free word ordering of Urdu language, this spatial predicate can change its position in the sentence which is not discussed by Butt et. al. However, in doctoral thesis of Raza [25], two postulates have been developed. One is the same as narrated by Butt et. al but the other one is related to position of a spatial predicate if it is moved in the beginning of the sentence and then it is considered to be a modifier MODF and not a PRD. The example of the spatial copula is given in 9.

```
                                              "woh laRkA mErA bHAI hE"

 CS 1:          S:41              ┌PRED  'hE<[3:laRkA], [7:bHAI]>'                          ┐
                                  │         ┌PRED  'laRkA'                              ┐
                                  │         │     ┌ ┌PRED  'woh'                     ┐ ┐ │
      ┌─────────┴────┐           │SUB   3 │SPEC │1│                                │ │ │
  NPnom:23   NPnom:35  VCMAIN:39  │         │     │2│CASE NOM, PERS 3, TYPE demonstrative-pronoun│ │ │
                                  │         │     └ └                              ┘ ┘ │
  ┌──┴──┐     ┌──┴──┐      │      │      23│CASE NOM, GEND masc, NUM sg, PERS 3, TYPE count-noun│
 PDEM:2 N:4 PPOSS:6 N:8  VCOP:11  │         └PRED  'bHAI'                                ┐
                                  │         ┌PRED  'bHAI'                             ┐ │
   │     │     │     │      │      │       7│     ┌ ┌PRED  'mErA'                   ┐ ┐ │ │
 woh:1 laRkA:3 mErA:5 bHAI:7 hE:9 │PRD   9 │SPEC │5│                              │ │ │ │
                                  │      11│     │6│GEND masc, NUM sg, TYPE possessive-pronoun│ │ │
                                  │      39 └     └                              ┘ ┘ │ │
                                  │         8│                                       │ │
                                  │      35│CASE NOM, GEND masc, NUM sg, PERS 3, TYPE count-noun│
                                  └41│GEND masc, MOOD indicative, NUM sg, TENSE pres        ┘
```
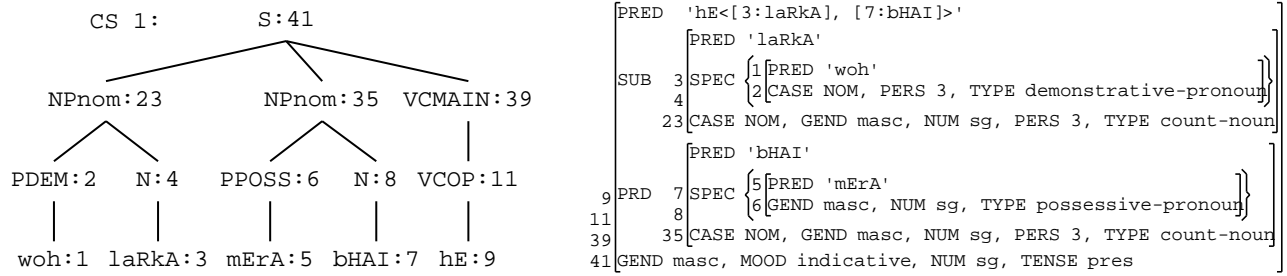
Fig. 4. The structure analysis of example 4

```
                                              "mErI laRkIyAN lambI tHIN"

 CS 1:          S:39              ┌PRED  'tHIN<[3:laRkIyAN], [5:lambI]>'                     ┐
                                  │         ┌PRED  'laRkIyAN'                           ┐
      ┌─────────┴────┐           │         │     ┌ ┌PRED  'mErI'                   ┐ ┐ │
 NPnom:21   ADJP:32  VCMAIN:37    │SUB   3 │SPEC │1│                              │ │ │
                                  │       4│     │2│GEND fem, TYPE possessive-pronoun│ │ │
  ┌──┴──┐      │         │        │         │     └ └                              ┘ ┘ │
 PPOSS:2 N:4  ADJ:6   VCOP:9      │      21│CASE NOM, GEND fem, NUM pl, PERS 3, TYPE count-noun│
                                  │       7│     ┌PRED  'lambI'              ┐
   │     │      │         │        │PRD   9│     6│GEND fem, TYPE adjective  │
 mErI:1 laRkIyAN:3 lambI:5 tHIN:7 │      37│    32└                          ┘
                                  └39│GEND fem, MOOD indicative, NUM pl, TENSE past         ┘
```
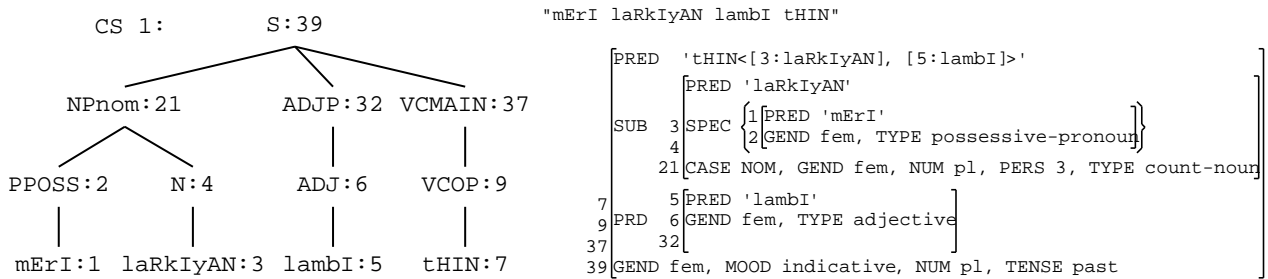
Fig. 5. The structure analysis of example 7

(9) kItAb   mEz=par   tHI.
Book.fem.3sg table.fem.3sg=SPT.on be.cop.past.fem.sg

'The book was on the table.'

If *mEz par* is shifted to the beginning of the sentence, then the readings of the sentence is totally changed and the translation of whole sentence will become like 'There was a book on the table' which is the case of an existential copula and it is also encoded in the grammar. The behavior of this existential copula verb is intransitive and a spatial MODF is normally at the position of a subject in an existential copular construction [18]. The c & f structures of example 9 are given in figure 7, in which the SUB *kItAb* agrees with the copula verb *tHI* in gender and number. The c & f structures of Freeze's theory are given in figure 8. Both the solution are incomplete and the only deficiency in this solution is that the PRD should be XCOMP-PRD and then *mEz* (table) should be the object of *par* (on) with the same SUB as discussed in section *3.2.3*. The classifier is deficient in this perspective, which will be dealt in future.

Spatial ad-positions are also possible in this construction and a sentence is presented in example 10. The treatment of this sentence having a spatial post-positional predicate is same like figure 7 and the agreement between SUB and copula verb is also the same. The possible feature rule for a spatial copular construction is similar to appendix A. All of the tense forms can be used with this construction without any effect on argument structure.

(10) kItAb   mEz=kE   nIcE
Book.fem.3sg table.fem.3sg=POSS under
tHI.
be.cop.past.fem.sg

'The book was under the table.'

3.2.5 Possessive Copula. Spatial case marker can also be used to identify an abstract location and then this abstract location will become the subject of a clause [22]. The case marker *par* (on) can be used to identify an abstract location of *z2ImahdArI* (responsibility) and *bOjH* (burden) as in example sentence 11. The case marker *mEN* (in) can be used to identify an abstract location of some inherent property like *mAmtA* (affection) and *jurAt* (courage) as in example sentence 12.

(11) nIdA=par   ta2lIm=kI
Nida.fem.sg=POSS.on education.fem.sg=POSS
z2ImahdArI   hE.
responsibility.fem.sg be.cop.pres.masc+fem.sg

'Nida has the responsibility of education.'

(12) Ali=mEN   bahut jurat
Ali.masc.sg=POSS.in great courage.fem.sg
hE.
be.cop.pres.masc+fem.sg

'Ali has a great deal of courage.'

The general reading of both the case markers *par* and *mEN* is spatial but these can interpret other readings as well, which can be seen in c & f structures of example 11 in figure 9. The subject *nIdA* has a case marker *par* and the copula verb agrees with PRD *z2ImahdArI* in number and gender. If SUB and PRD positions are interchanged with each other then SUB becomes PRD and PRD becomes SUB. Therefore, two readings including a possession 'has' and a spatial 'on' are possible. The sentence can be translated as *Nida has the responsibility of education* or *The responsibility of education is on Nida*. Both are semantically correct alternate to each other. The c & f structures of example 12 are not provided here due to limited length of the article.

```
                                          "nIdA sOlah sAl kI hE"

                                    ⌈PRED  'hE<[1:nIdA], [7:kI]>'                                      ⌉
                                    |      ⌈PRED 'nIdA'                                              ⌉  |
      CS 1:      S:50              |SUB  2|CASE NOM, GEND fem, NUM sg, PERS 3, TYPE proper-noun      |  |
                                    |   24⌊                                                          ⌋  |
                                    |       ⌈PRED 'kI'                                             ⌉   |
NPnom:24   KPattr:44  VCMAIN:48     |       |        ⌈PRED 'sAl'                               ⌉⌉   |
                                    |       |        |     ⌈PRED 'sOlah'                      ⌉|⌉   |
   |        /\         |            |PRD NOM{ 5|MODF 3|4|GEND masc, NUM pl, TYPE cardinal-quantifier||⌋   |
NPROP:2 NPnom:38 CM:8 VCOP:11       |    7|       6⌊    ⌊                                   ⌋|⌋   |
                                    |  9  |     8 38|CASE NOM, GEND masc, NUM sg, PERS 3, TYPE temporal-noun|   |
   |     /\     |      |            | 11  ⌊                                                     ⌋   |
 nIdA:1 Q:4   N:6   kI:7  hE:9      | 48   44⌊CASE POSS, GEND fem                                ⌋   |
                                    | 50⌊GEND fem, MOOD indicative, NUM sg, TENSE pres              ⌋
         |     |
      sOlah:3 sAl:5
```
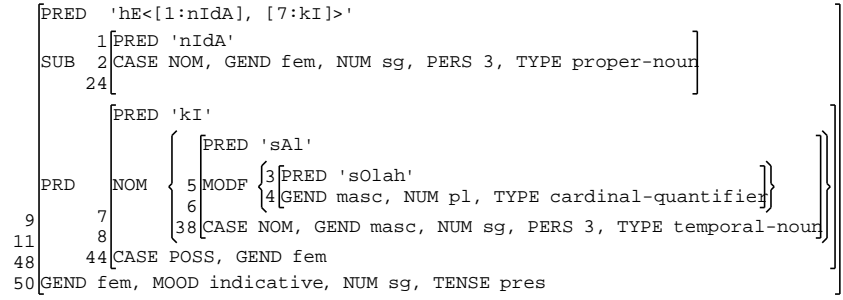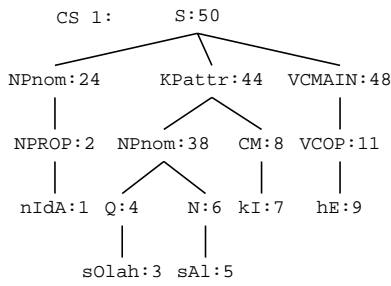
Fig. 6.   The structure analysis of example 8

```
                                          "kItAb mEz par tHI"

                                    ⌈PRED  'tHI<[1:kItAb], [5:par]>'                             ⌉
      CS 1:      S:42              |      ⌈PRED 'kItAb'                                        ⌉ |
                                    |SUB  2|CASE NOM, GEND fem, NUM sg, PERS 3, TYPE count-noun | |
                                    |   23⌊                                                     ⌋ |
NPnom:23   KPspt:44  VCMAIN:40      |       ⌈PRED 'par'                                       ⌉ |
                                    |       |     ⌈PRED 'mEz'                              ⌉⌉ |
   |        /\         |            | PRD 5|NOM{ 3|4|CASE NOM, GEND fem, NUM sg, PERS 3, TYPE spatial-noun||⌋ |
  N:2      N:4  CM:6  VCOP:9        | 7   |      6⌊    ⌊                                 ⌋⌋ |
                                    | 9    44⌊CASE SPT                                       ⌋ |
   |        |    |      |           | 40⌊                                                     ⌋
kItAb:1  mEz:3 par:5  tHI:7         | 42⌊GEND fem, MOOD indicative, NUM sg, TENSE past           ⌋
```
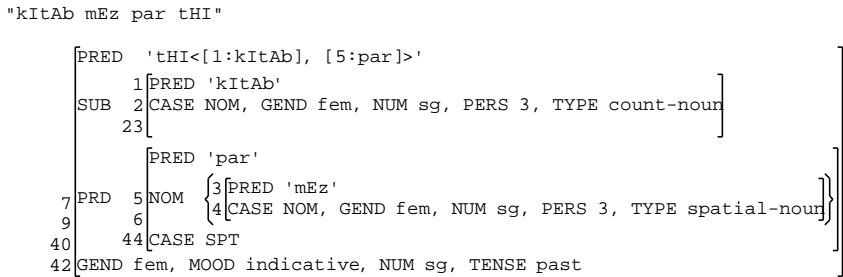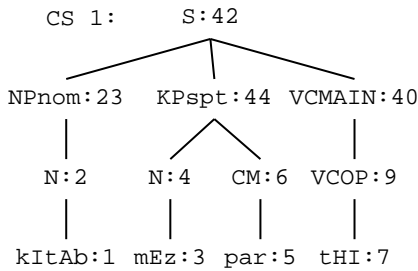
Fig. 7.   The structure analysis of example 9

An abstract location can be identified through a spatial case marker (as discussed) and also through dative case marker. A dative case marker [9] can also be used to identify experiencer/SUB having some experience/PRD in a copular construction as given in example sentence 13. The respective c & f structures are depicted in figure 10. According to these structures, an agreement is lying between PRD *buxAr* and copula verb *hE*.

(13)   Ali=kO   buxAr   hE.
       Ali.masc.sg=DAT buxAr.masc be.cop.pres.masc+fem.sg
       'Ali has a fever/ Ali suffers from fever.'

A copula verb also encodes the possession relation 'has/have' of Urdu possessive case markers *ka/kI/kE* like in English. The example sentences are given in 14 and 15. According to Mohanan [22], a copula verb agrees with unmarked possessees and if it is not present then a default masculine singular form is used. In a sentence, an argument having a possessive case marker acts as a subject SUB.

(14)   Ali=kI   Ek behan
       Ali.masc.sg=POSS one behan.fem.sg
       hE.
       be.cop.pres.masc+fem.sg
       'Ali has one sister.'

(15)   woh kItAb   Ali=kI
       That book.fem.sg Ali.masc.sg=POSS
       hE.
       be.cop.pres.masc+fem.sg
       'That book is of Ali.'

There is a spatial postposition *pAs* (near) used to encode the possession 'has/have' relationship with in the copular construction as

discussed in examples 11 and 12. A nominal having a *pAs* spatial postposition is considered as the subject of a sentence [22]. McGregor identified genitives *kA/kI/kE* in the category of permanent possession and *pAs* construction in the category of contingent possession in [21]. Similarly, in [23], Pandharipande argued that genitive is used to express emotional attachment or intimacy, while *pAs* construction is used to express material possession. Later on, Sebastian Sulger [27] categorized these two constructions as individual level and stage level predication after performing a test given in [13, 17, 20]. The related sentences to this concept of discussion are given in example 16 and 17.

(16)   Ali=kE   pAs sOnE=kI   angUTHI
       Ali.masc.sg=POSS near gold.masc.obl=POSS ring.fem.sg
       hE.
       be.cop.pres.masc+fem.sg
       'Ali has a ring of gold.'

(17)   mErI kItAb   nIdA=kE   pAs
       My   book.fem.sg nIda.fem.sg=POSS near
       hE.
       be.cop.pres.masc+fem.sg
       'My book is with Nida.'

Both of the examples have a *pAs* (near) spatial postposition in it. In example 16, the reading is possessive and a making of KPposs is depicted in the c structure of figure 11. In example 17, the owner of the book is someone else but it is temporarily shifted to another abstract location *Nida*. Its translation is changed due to a spatial postposition with an argument of a sentence. In example 17, 'My book' is a subject and 'with Ali' is a predicate link which is not displayed here. The possible feature rule with all options is given in
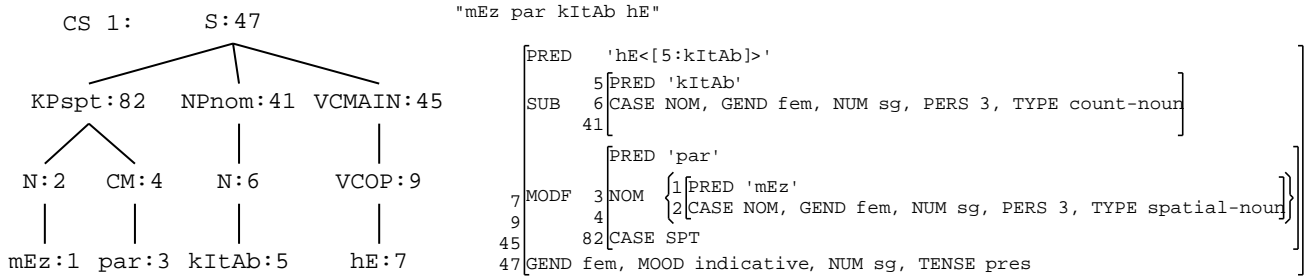
```
CS 1:        S:47              "mEz par kItAb hE"

                                    ┌PRED  'hE<[5:kItAb]>'
KPspt:82  NPnom:41  VCMAIN:45       │         5┌PRED 'kItAb'
                                    │SUB      6│CASE NOM, GEND fem, NUM sg, PERS 3, TYPE count-noun
                                    │        41└
 ┌──┐      │         │              │           ┌PRED 'par'
 N:2 CM:4  N:6     VCOP:9           │         3 │NOM ⎧1┌PRED 'mEz'
                                    │7 MODF     │    ⎨ │
 │    │    │         │              │9        4 │    ⎩2└CASE NOM, GEND fem, NUM sg, PERS 3, TYPE spatial-noun
                                    │45        82└CASE SPT
mEz:1 par:3 kItAb:5  hE:7           47└GEND fem, MOOD indicative, NUM sg, TENSE pres
```

Fig. 8.   The structure analysis of Freeze's Theory[18] for Urdu existential copular construction

```
              CS 1:              S:51

       KPposs:29       NPnom:62         VCMAIN:66

     ┌─────┐        ┌───────────┐           │
  NPROP:2  CM:4  KPposs:55    N:10       VCOP:13

     │      │     ┌─────┐       │            │
  nIdA:1  par:3 N:6   CM:8  z2ImahdArI:9  hE:11

                 │     │
              ta2lIm:5 kI:7
```

"nIdA par ta2lIm kI z2ImahdArI hE"

```
┌PRED  'hE<[3:par], [9:z2ImahdArI]>'
│         ┌PRED 'par'
│         │
│SUB   3 │NOM ⎧1┌PRED 'nIdA'
│      4 │    ⎨ │
│     29 │    ⎩2└CASE NOM, GEND fem, NUM sg, PERS 3, TYPE proper-noun
│         └CASE POSS
│         ┌PRED 'z2ImahdArI'
│         │         ⎧ ┌PRED 'kI'
│         │         ⎪ │
│PRD      │SPEC ⎨ 7 │NOM ⎧5┌PRED 'ta2lIm'
│  11   9 │         ⎪ 8 │    ⎨ │
│  13  10 │         ⎩55 │    ⎩6└CASE NOM, GEND fem, NUM sg, PERS 3, TYPE common-noun
│  66     │         55└CASE POSS, GEND fem
│      62 └CASE NOM, GEND fem, NUM sg, PERS 3, TYPE common-noun
51└GEND fem, MOOD indicative, NUM sg, TENSE pres
```
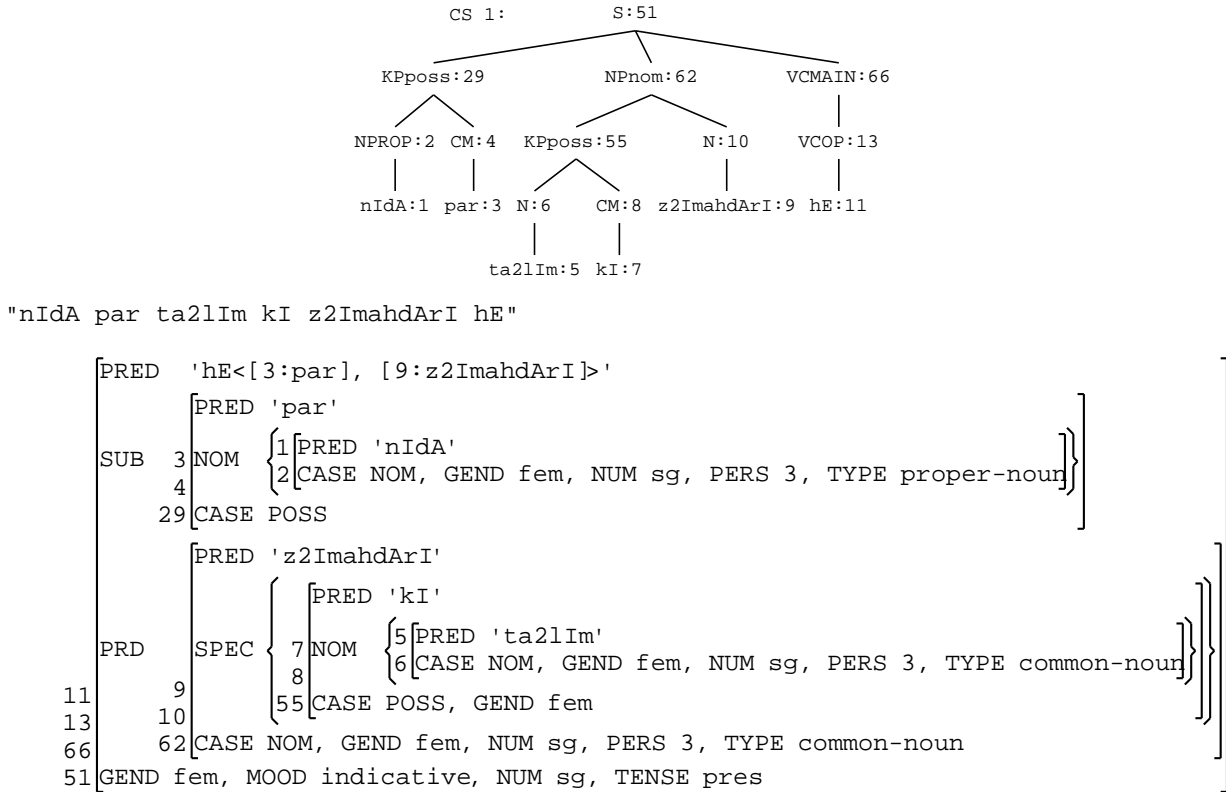
Fig. 9.   The structure analysis of example 11

appendix D. The agreement in both examples is with that argument which does not have a case marker. An agreement of a PRD and a VCOP is existed in first sentence and an agreement of a SUB and a VCOP is existed in second sentence.

3.2.6 Tense Auxiliary & Modal Copula. The non aspectual form of verb *hO* can also be used as a tense auxiliary. The present tense examples are similar to example 1 and 2 given in section 3.1. If the copula verb *hO* is replaced with some other verb like *karnA* (to do) then the tense auxiliary will behave as a copula verb. Similarly, after a same modification, if a tense auxiliary is changed to past tense auxiliary like *tHA, tHI, tHE , tHIN* then it will be the case of copula verb *hO* with past tense. The case of future is different because it provides two readings. First is a stative modal reading and the sec-

ond is a dynamic reading. When the future tense is used in a sentence then *gA, gI, gE, gIN* will be the future tense auxiliary and its behavior is not as a copula verb. The copular construction in future is possible but then it requires the verb *hO* in subjunctive/root form. Similarly, in passive construction, it requires subjunctive form of *jA* with the verb *hO*. A sentence having a future tense is given in example 18. The passive construction normally provides a dynamic reading and the other provides a stative modal reading. A tense auxiliary is an obligatory source for provision of tense and copular construction except the future tense [11]. The syntactic feature rule is same as in appendix D except an adjective phrase AP is also possible.
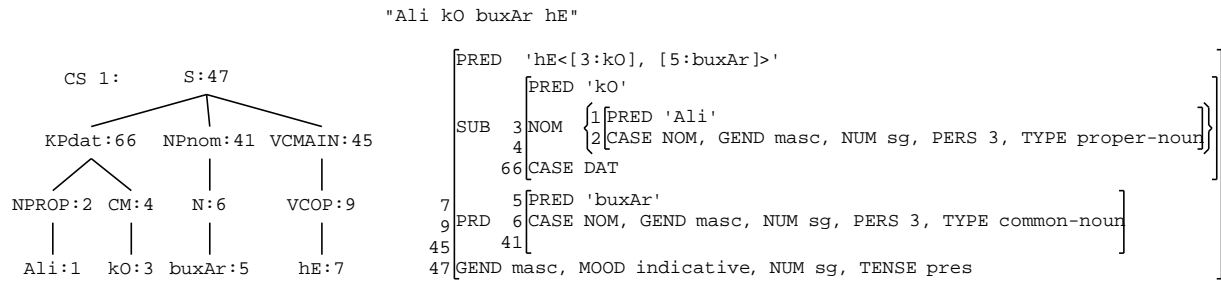
```
                                         "Ali kO buxAr hE"

                                              PRED  'hE<[3:kO], [5:buxAr]>'
       CS 1:        S:47                             PRED 'kO'
                                                           1 PRED 'Ali'
                                              SUB  3 NOM
                                                           2 CASE NOM, GEND masc, NUM sg, PERS 3, TYPE proper-noun
   KPdat:66  NPnom:41  VCMAIN:45                   4
                                                66 CASE DAT
                                               7     5 PRED 'buxAr'
 NPROP:2  CM:4   N:6      VCOP:9                9 PRD 6 CASE NOM, GEND masc, NUM sg, PERS 3, TYPE common-noun
                                              45    41
   Ali:1  kO:3  buxAr:5   hE:7                 47 GEND masc, MOOD indicative, NUM sg, TENSE pres
```

Fig. 10.   The structure analysis of example 13

```
                         CS 1:                 S:65

                 PPposs:45            NPnom:76           VCMAIN:80

           KPposs:31  POSTP:6   KPposs:69      N:12       VCOP:15

        NPROP:2  CM:4   pAs:5    N:8   CM:10  angUTHI:11   hE:13

          Ali:1  kE:3           sOnE:7  kI:9
```

"Ali kE pAs sOnE kI angUTHI hE"

```
PRED  'hE<[5:pAs], [11:angUTHI]>'
        PRED 'pAs'
                      PRED 'kE'
                            1 PRED 'Ali'
SUB    SPEC   3 NOM
              4           2 CASE NOM, GEND masc, NUM sg, PERS 3, TYPE proper-noun
        5    31 CASE POSS, GEND masc
        6
       45 TYPE possessive-postposition
        PRED 'angUTHI'
                      PRED 'kI'
                            7 PRED 'sOnE'
PRD    SPEC   9 NOM
             10           8 CASE OBL, GEND masc, NUM sg, PERS 3, TYPE common-noun
       11   69 CASE POSS, GEND fem
13     12
15     76 CASE NOM, GEND fem, NUM sg, PERS 3, TYPE count-noun
80
65 GEND fem, MOOD indicative, NUM sg, TENSE pres
```

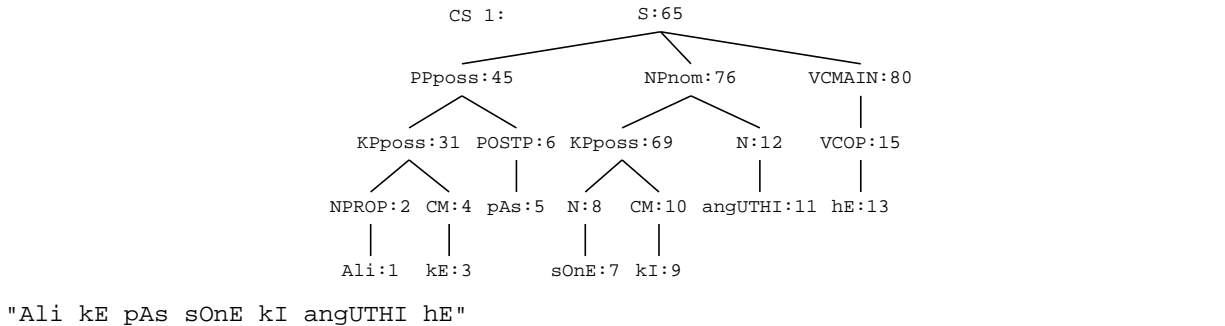Fig. 11.   The structure analysis of example 16

(18) mErE kAm   acHE   hoN
      My   work.masc.pl fine.masc.pl become.cop.subjunctive
      gE   .
      be.fut.masc.pl

      'My works will become fine.'

The c & f structures of example 18 are presented in figure 12. The VCMAIN agrees with the SUB and PRD both at a time. The agreement effect can be seen in the glossing of future tense auxiliary *gE*, which contains a future tense, a masculine gender and a plural number in order.

3.3   Aspectual Copula

Aspectual forms of copula verb *hO* (be/become) can be divided into imperfective and perfective forms as follows.

3.3.1   Imperfective Copula.   Imperfective form is used for generic reading [12] and in Urdu imperfective form of verb *hO* is also used for similar purpose. The example sentences are given in 19 and 20 as follows.
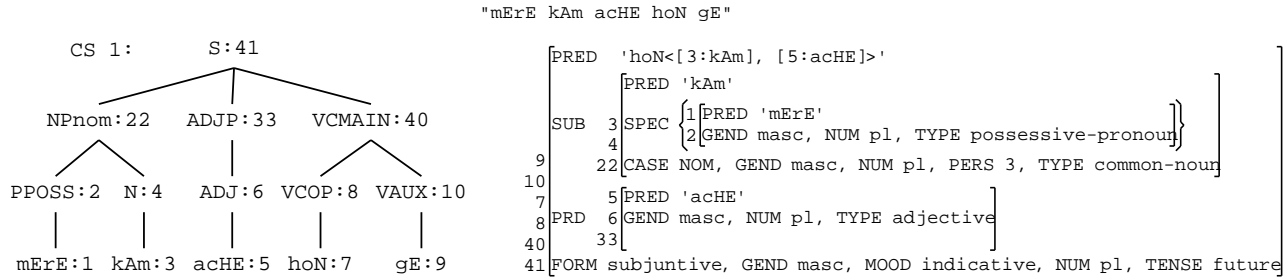
```
                                   "mErE kAm acHE hoN gE"

    CS 1:        S:41                PRED  'hoN<[3:kAm], [5:acHE]>'
                                           PRED 'kAm'
                                                     1 PRED 'mErE'
   NPnom:22  ADJP:33  VCMAIN:40      SUB  3 SPEC      2 GEND masc, NUM pl, TYPE possessive-pronoun
                                        9
                                       10  22 CASE NOM, GEND masc, NUM pl, PERS 3, TYPE common-noun
   PPOSS:2  N:4    ADJ:6  VCOP:8 VAUX:10   7  5 PRED 'acHE'
                                        8 PRD 6 GEND masc, NUM pl, TYPE adjective
                                       40    33
   mErE:1  kAm:3  acHE:5  hoN:7    gE:9 41 FORM subjuntive, GEND masc, MOOD indicative, NUM pl, TENSE future
```

Fig. 12. The structure analysis of example 18

(19) Ali   sargOdHA=mEN hotA
      Ali.masc.sg sargodha=SPT.in be.cop.imperf.masc.sg
      hE.
      be.pres.masc+fem.sg
      'Ali is in Sargodha.'

(20) zIndagI   Ek na2mat   hotI
      Life.fem.sg a   blessing.fem.sg be.cop.imperf.fem.sg
      hE.
      be.pres.masc+fem.sg
      'Life is a blessing.'

In both of the examples, the imperfective forms of *hotA* and *hotI* of verb *hO* are used to build a generic/habitual relation as *Ali* lives in *Sargodha* and *Life* is a *blessing*. A common noun in Urdu can be indefinite or specific depending on the context [15]. In example 19, *Ali* is not a common noun, so it is specific to that person only. However, in example 20, *Life* is a common noun and it is applied on the whole class of humans and is an indefinite. The c and f structures of example 20 are given in figure 13 and the rest of the sentences are not displayed due to page limit of this article. Here, 'Life's is the SUB of the sentence and 'Blessing' is the PRD of the sentence. In Urdu, if there is no case marker with the SUB, then the verb agrees with the SUB in number and gender.

3.3.2 Perfective Copula. When a perfective form of verb *hO* is used, then only dynamic reading is possible with the meaning of become/happen. The different perfective forms *hUA, hUE, hUI, hUIN* are possible. Some of the example sentences can be seen in 21, and 22. The c & f structures of example sentence 21 are displayed in figure 14. *Ali* is the SUB and *xUS* is the PRD of the sentence. Since, there is no any case marker *kA/kI/kE/kO/mEN/sE* with the SUB, hence the verb agrees with the SUB in number and gender in this sentence. Same is the case with a sentence given in example 22.

(21) Ali   kal   bahut xUS   huA.
      Ali.masc.sg yesterday much happy be.cop.perf.masc.sg
      'Ali became much happy yesterday.'

(22) mErE bAl   lambE   huE.
      My   hairs.masc.pl long.masc.pl be.cop.perf.masc.pl
      'My hairs became long.'

The appendix D is the almost syntactic feature rule for this perfective and imperfective form discussed.

4. CONCLUSION

The classifier concludes eleven different classifications of Urdu copula verb *hO* (be/become) based on linguistic theories and the native knowledge of Urdu speakers. These classifications can be divided into further categories and can be evaluated accordingly. The future work of identification and classification includes a verbal copula, a particle copula, a zero copula and an inflectional copula. The spatial copula can contain the 'part to whole' concept with a new category. The possessum case of taking position of a SUB while having a position of PRD at a time is not considered in the possessive copula classification dicussed. A word *apnA* (own) is also not discussed in section *3.2.5*. The verb *hO* as a modal is not discussed with some other possible examples. A number of issues regarding copular construction in Urdu are still remaining to explore like the word *nahIN* (not) in absence of a tense auxiliary and the verb *hO* as an emphasis on 'being'. Only noun-verb and verb-verb complex predicates are discussed and the test suite of the classifier does not contain any example of adjective-verb complex predicates. The classifier lacks in light verb discussion related to copular construction. However, the classifier implemented is still a state-of-the-art work for a under resource language Urdu and the remaining issues will be attempted soon.

5. REFERENCES

[1] Qaiser Abbas. Building a hierarchical annotated corpus of urdu: the URDU.KON-TB treebank. In *Computational Linguistics and Intelligent Text Processing*, pages 66–79. Springer, 2012.

[2] Qaiser Abbas, Nayyara Karamat, and Sadia Niazi. Development of tree-bank based probabilistic grammar for urdu language. *International Journal of Electrical & Computer Science*, 9(09):231–235, 2009.

[3] Qaiser Abbas and A Nabi Khan. Lexical functional grammar for urdu modal verbs. In *Emerging Technologies, 2009. ICET 2009. International Conference on*, pages 7–12. IEEE, 2009.

[4] Collin F Baker, Charles J Fillmore, and John B Lowe. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics, 1998.

[5] Miriam Butt. Machine translation and complex predicates. In *Conference for processing nat urlicher language (KONVENS 94) Osterreichische society f ur AI*, pages 62–71, 1994.

[6] Miriam Butt. The light verb jungle. In *Workshop on Multi-Verb Constructions*, 2003.

[7] Miriam Butt. 3 the light verb jungle: still hacking away. *Complex predicates: cross-linguistic perspectives on event structure*, page 48, 2010.

```
                        "zIndagI Ek na2mat hotI hE"

                                      PRED 'hotI<[1:zIndagI], [5:na2mat]>'
   CS 1:          S:45                     1 PRED 'zIndagI'
                                      SUB  2 CASE NOM, GEND fem, NUM sg, PERS 3, TYPE common-noun
 NPnom:25 NPnom:38    VCMAIN:44          25
                                       9        PRED 'na2mat'
   |      /\          /\               10        MODF {3 PRED 'Ek'
  N:2   Q:4  N:6  VCOP:8 VAUX:10        7  5          {4 GEND masc, NUM sg, TYPE cardinal-quantifier}
   |     |    |     |     |     PRD       6
zIndagI:1 Ek:3 na2mat:5 hotI:7 hE:9    44 38 CASE NOM, GEND fem, NUM sg, PERS 3, TYPE count-noun
                                       45 FORM imperfective, GEND fem, MOOD indicative, NUM sg, TENSE pres
```
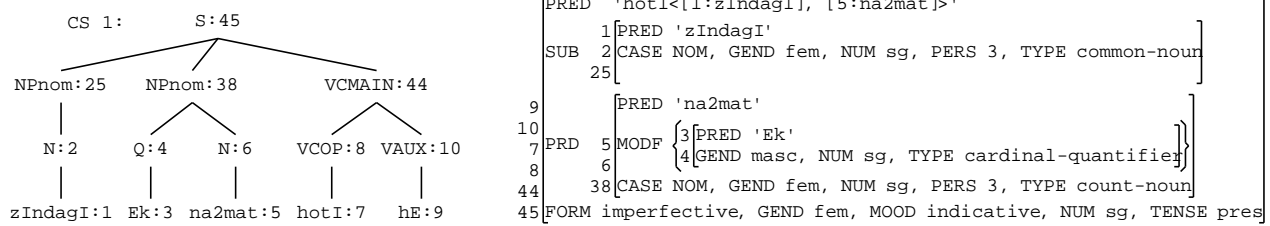
Fig. 13. The structure analysis of example 20

```
                        "Ali kal bahut xUS huA"

                                      PRED 'huA<[1:Ali], [7:xUS]>'
                                           1 PRED 'Ali'
                                      SUB  2 CASE NOM, GEND masc, NUM sg, PERS 3, TYPE proper-noun
   CS 1:          S:61                   23
                                               PRED 'xUS'
 NPnom:23 NPmod:43 ADJP:54  VCMAIN:59   PRD  7 MODF {5 PRED 'bahut'
                                             8        {6 TYPE degree-adverb}
 NPROP:2   N:4  ADV:6 ADJ:8  VCOP:10        54 TYPE adjective
   |        |     |    |       |             9        {3 PRED 'kal'
 Ali:1    kal:3 bahut:5 xUS:7 huA:9     MODF 10       {4 GEND masc, NUM sg, PERS 3, TYPE temporal-noun}
                                             59   43
                                        61 FORM perfective, GEND masc, MOOD indicative, NUM sg, TENSE past
```
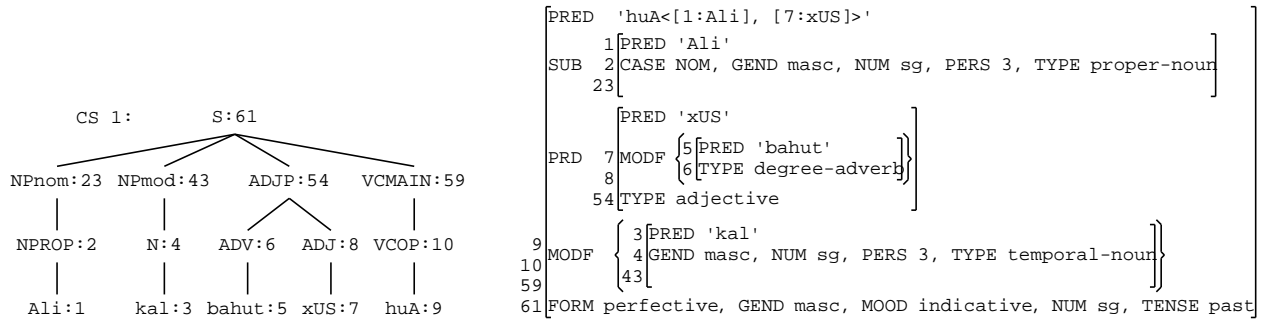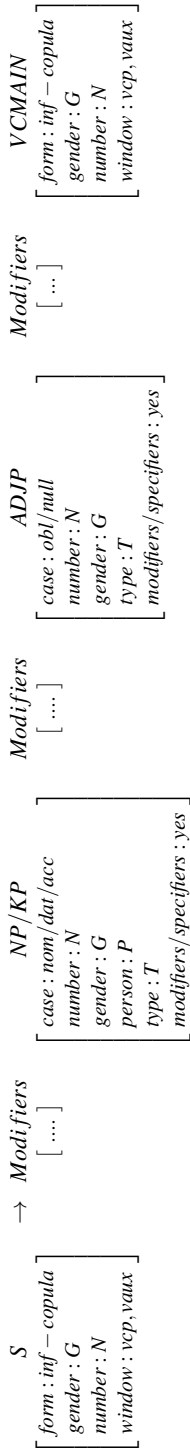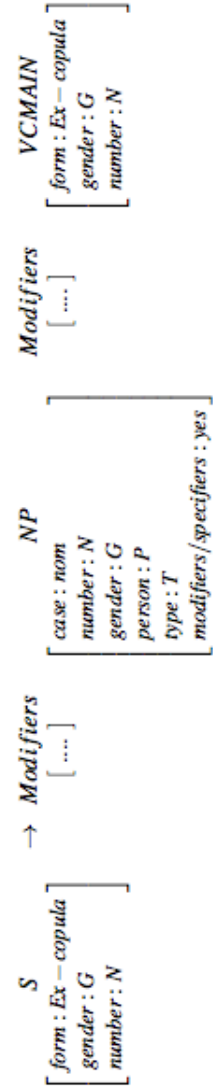
Fig. 14. The structure analysis of example 21

[8] Miriam Butt, Tracy King, Frédérique Segond, and Maria-Eugenia Nino. *A grammar writer's cookbook*, volume 95. CSLI publications Stanford, 1999.

[9] Miriam Butt and Tracy Holloway King. The status of case. In *Clause structure in South Asian languages*, pages 153–198. Springer, 2004.

[10] Miriam Butt and Gillian Ramchand. Complex aspectual structure in hindi/urdu. *M. Liakata, B. Jensen, & D. Maillat, Eds*, pages 1–30, 2001.

[11] Miriam Butt and Jafar Rizvi. Tense and aspect in urdu. *Layers of Aspect. Stanford: CSLI Publications*, 2010.

[12] Gregory N Carlson and Francis Jeffry Pelletier. *The generic book*. University of Chicago Press, 1995.

[13] Gregory Norman Carlson. Reference to kinds in english. 1977.

[14] Timothy Jowan Curnow. Towards a cross-linguistic typology of copula constructions. In *the 1999 Conference of the Australian Linguistic Society*, pages 1–9, 2000.

[15] Veneeta Dayal. The singular-plural distinction in hindi generics. In *Proceedings of SALT II*, pages 39–58, 1992.

[16] Renaat Declerck. *Studies on copular sentences, clefts and pseudo-clefts*, volume 5. Walter de Gruyter, 1988.

[17] Molly Diesing. Bare plural subjects and the stage/individual contrast. In *Genericity in Natural Language: Proceedings of the 1988 Tübingen Conference*, pages 107–154, 1988.

[18] Ray Freeze. Existentials and other locatives. *Language*, pages 553–595, 1992.

[19] Kees Hengeveld. *Non-verbal predication: theory, typology, diachrony*, volume 15. Walter de Gruyter, 1992.

[20] Angelika Kratzer. Stage-level and individual-level predicates. *The generic book*, 125:175, 1995.

[21] Ronald Stuart McGregor. *Outline of Hindi grammar, with exercises*. Oxford University Press, 1995.

[22] Tara Mohanan. *Argument structure in Hindi*. Center for the Study of Language (CSLI), 1994.

[23] Rajeshwari Pandharipande. Interface of lexicon and grammar: Some problems in hindi grammar in dimensions of south asian linguistics. *Studies in the Linguistic Sciences Urbana, Ill.*, 11(2):77–100, 1981.

[24] Ghulam Raza. Analyzing the structure of urdu nps with multiple genitives. In *Proceedings of the Conference on Language and Technology*, 2010.

[25] Ghulam Raza. Subcategorization acquisition and classes of predication in urdu. *Institutional Repository of University of Konstanz (KOPS)*, 2011.

[26] Karin Kipper Schuler. Verbnet: A broad-coverage, comprehensive verb lexicon. 2005.

[27] Sebastian Sulger, Miriam Butt, and Tracy Holloway King. A parallel analysis of have-type copular constructions in two have-less indo-european languages. In *Proceedings of 16th LFG conference held in Hong Kong, CSLI Publications*, volume 112, 2011.

APPENDIX

A.

$$S \left[\begin{array}{l} form : inf - copula \\ gender : G \\ number : N \\ window : vcp, vaux \end{array} \right] \rightarrow Modifiers \left[.... \right] \; NP/KP \left[\begin{array}{l} case : nom/dat/acc \\ number : N \\ gender : G \\ person : P \\ type : T \\ modifiers/specifiers : yes \end{array} \right] \; Modifiers \left[.... \right] \; ADJP \left[\begin{array}{l} case : obl/null \\ number : N \\ gender : G \\ type : T \\ modifiers/specifiers : yes \end{array} \right] \; Modifiers \left[... \right] \; VCMAIN \left[\begin{array}{l} form : inf - copula \\ gender : G \\ number : N \\ window : vcp, vaux \end{array} \right]$$

B.

$$S \left[\begin{array}{l} form : Ex - copula \\ gender : G \\ number : N \end{array} \right] \rightarrow Modifiers \left[... \right] \; NP \left[\begin{array}{l} case : nom \\ number : N \\ gender : G \\ person : P \\ type : T \\ modifiers/specifiers : yes \end{array} \right] \; Modifiers \left[... \right] \; VCMAIN \left[\begin{array}{l} form : Ex - copula \\ gender : G \\ number : N \end{array} \right]$$

C.

$$S \begin{bmatrix} form: Attr-copula \\ gender: G \\ number: N \\ Tense: All \\ window: prog, vaux \end{bmatrix} \rightarrow Modifiers [....] \quad NP \begin{bmatrix} case: nom \\ number: N \\ gender: G \\ person: P \\ type: T \\ modifiers/specifiers: yes \end{bmatrix} \quad KP \begin{bmatrix} case: poss \\ number: N \\ gender: G \\ person: P \\ type: T \\ modifiers/specifiers: yes \end{bmatrix} \quad VCMAIN \begin{bmatrix} form: Attr-copula \\ gender: G \\ number: N \\ Tense: All \\ window: prog, vaux \end{bmatrix}$$

D.

$$S \begin{bmatrix} form: Poss-copula \\ gender: G \\ number: N \\ Tense: All \end{bmatrix} \rightarrow Modifiers [....] \quad KP/NP \begin{bmatrix} case: poss/nom/dat \\ number: N \\ gender: G \\ person: P \\ type: T \\ modifiers/specifiers: yes \end{bmatrix} \quad Modifiers [....] \quad KP/NP \begin{bmatrix} case: nom/poss/ \\ number: N \\ gender: G \\ person: P \\ type: T \\ modifiers/specifiers: yes \end{bmatrix} \quad Modifiers [....] \quad VCMAIN \begin{bmatrix} form: Poss-copula \\ gender: G \\ number: N \\ Tense: All \end{bmatrix}$$