

Computational Grammar - 2

Tafseer Ahmed

Universal Dependency structures in CoNLL format
Noun / Adj + Verb Complex Predicate
Copula(r Verbs)

Background

Universal Dependencies

Universal Dependencies (UD) is a framework for consistent annotation of grammar (parts of speech, morphological features, and syntactic dependencies) across different human languages. UD is an open community effort with over 300 contributors producing more than 150 treebanks in 90 languages. If you're new to UD, you should

Universal Dependencies

	Nominals	Clauses	Modifier words	Function Words
Core arguments	<u>nsubj</u> <u>obj</u> <u>iobj</u>	<u>csubj</u> <u>ccomp</u> <u>xcomp</u>		
Non-core dependents	<u>obl</u> <u>vocative</u> <u>expl</u> <u>dislocated</u>	<u>advcl</u>	<u>advmod</u> * <u>discourse</u>	<u>aux</u> <u>cop</u> <u>mark</u>
Nominal dependents	<u>nmod</u> <u>appos</u> <u>nummod</u>	<u>acl</u>	<u>amod</u>	<u>det</u> <u>clf</u> <u>case</u>
Coordination	MWE	Loose	Special	Other
<u>conj</u> <u>cc</u>	<u>fixed</u> <u>flat</u> <u>compound</u>	<u>list</u> <u>parataxis</u>	<u>orphan</u> <u>goeswith</u> <u>reparandum</u>	<u>punct</u> <u>root</u> <u>dep</u>

UD Corpus of Urdu

→  github.com/UniversalDependencies/UD_Urdu-UDTB/tree/dev



dan-zeman committed 98e7a5e on May 1, 2019 ...



21 commits




2 branches



9 tags

 **.gitignore** removed test from gitignore

3 years ago

 **CONTRIBUTING.md** Updated CONTRIBUTING.md.

2 years ago

 **LICENSE.txt** Hindi --> Urdu.

3 years ago

 **README.md** Normalized Unicode and fixed some other errors.

14 months ago

 **stats.xml** Normalized Unicode and fixed some other errors.

14 months ago

 **ur_udtb-ud-dev.conllu** Normalized Unicode and fixed some other errors.

14 months ago

 **ur_udtb-ud-test.conllu** Normalized Unicode and fixed some other errors.

14 months ago

 **ur_udtb-ud-train.conllu** Normalized Unicode and fixed some other errors.

14 months ago

UD Corpus for Sindhi

github.com/UniversalDependencies/UD_Sindhi-MazharDootio/blob/dev/sd_mazhardootio-ud-test.conllu

Branch: dev UD_Sindhi-MazharDootio / sd_mazhardootio-ud-test.conllu Go to

dan-zeman CCONJ Latest commit a99bfc4 on Apr 23

1 contributor

8858 lines (8193 sloc) 671 KB Raw Blame

```
1 # sent_id = 1
2 # text = يقين ڪرڻ سان اڪثر ڏوسو ئي ملي ٿو .
3 1 يقين يقين NOUN NN_اسم Case=Nom|Form=Simple|Gender=Masc|Number=Sing|Person=3 0 dep
4 2 ڪرڻ ڪر VERB VB_فعل Form=Simple|Number=Sing 0 dep _ Translit=karan
5 3 سان سان ADP IN_حرف-جر Form=Bound|Number=Sing 0 dep _ Translit=Saana
6 4 اڪثر اڪثر ADV RB_ظرف Form=Simple|Number=Sing 0 dep _ Translit=Aksar
7 5 ڏوسو ڏوسو NOUN NN_اسم Case=Acc|Form=Simple|Gender=Masc|Number=Sing|Person=3 0 dep
8 6 ئي ئي ADV RB_ظرف Form=Simple|Number=Sing 0 dep _ Translit=Ee
9 7 ملي مل VERB VB_فعل Form=Bound|Number=Sing 0 dep _ Translit=Milay
10 8 ٿو ٿو VERB VB_فعل Form=Simple|Number=Sing 0 dep _ Translit=tho
11 9 . . PUNCT _-پورو دم _ 0 dep _ Translit=PooroDam
```

CoNLL Format

- CoNLL (Conference on Natural Language Learning) format
- Representing graph (and other tags) in text file

Id

Word

Lemma

Coarse Grained POS

Fine Grained POS

Features

Host

Dependency Type

CoNLL Format

dependency-conll - Notepad							
File	Edit	Format	View	Help			
1	ذبین	ذبین	Adj	Adj	-	2	amod
2	لڑکیاں	لڑکی	Noun	NN	-	6	subj
3	نے	نے	Adp	PP	-	2	case
4	اچھی	اچھا	Adj	Adj	-	5	amod
5	کتابیں	کتاب	Noun	NN	-	6	obj
6	پڑھیں	پڑھ	Verb	VB	-	0	ROOT
7	تھیں	ے	Aux	Aux	-	6	aux

Creating UD structures

لڑکا کتاب پڑھتا ہے۔
لڑکی نے کتاب پڑھی۔
لڑکے نے سبق یاد کیا۔
لڑکی نے میز صاف کی۔
لڑکی ذہین ہے۔
لڑکا گھر پر ہے۔
لڑکا وکیل بن گیا۔

لڑکا کتاب پڑھتا ہے -

<https://urd2.let.rug.nl/~kleiweg/conllu/>

← → ↻ urd2.let.rug.nl/~kleiweg/conllu/

Upload a file with one or more sentences annotated in [CoNLL-U](#) format:

Choose File

Submit

Here is an [example](#)

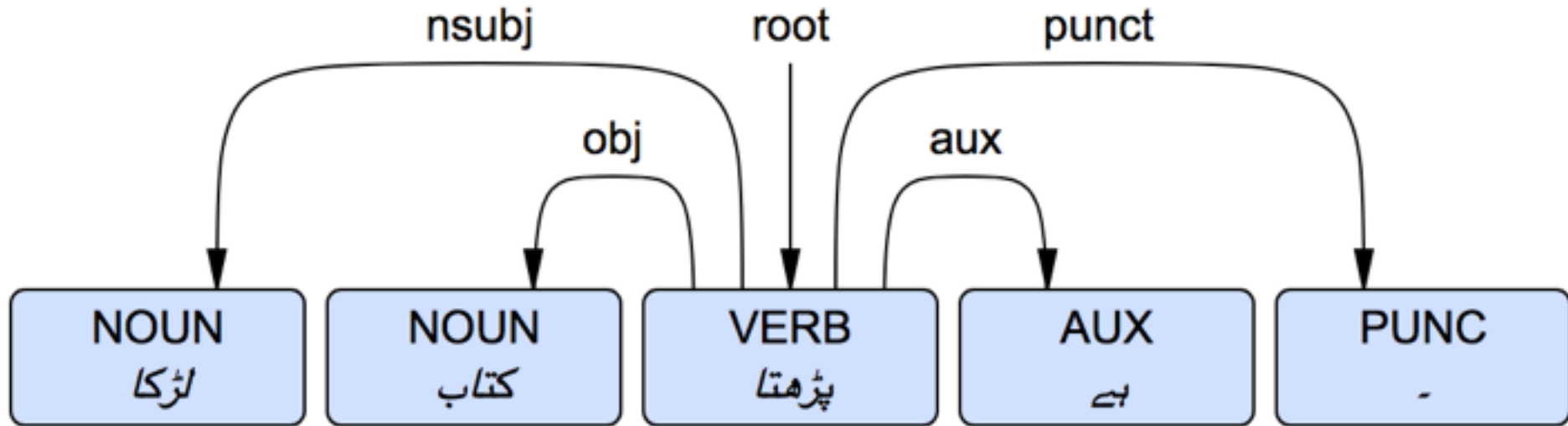
— — *OR* — —

Enter something in CoNLL-U format here:

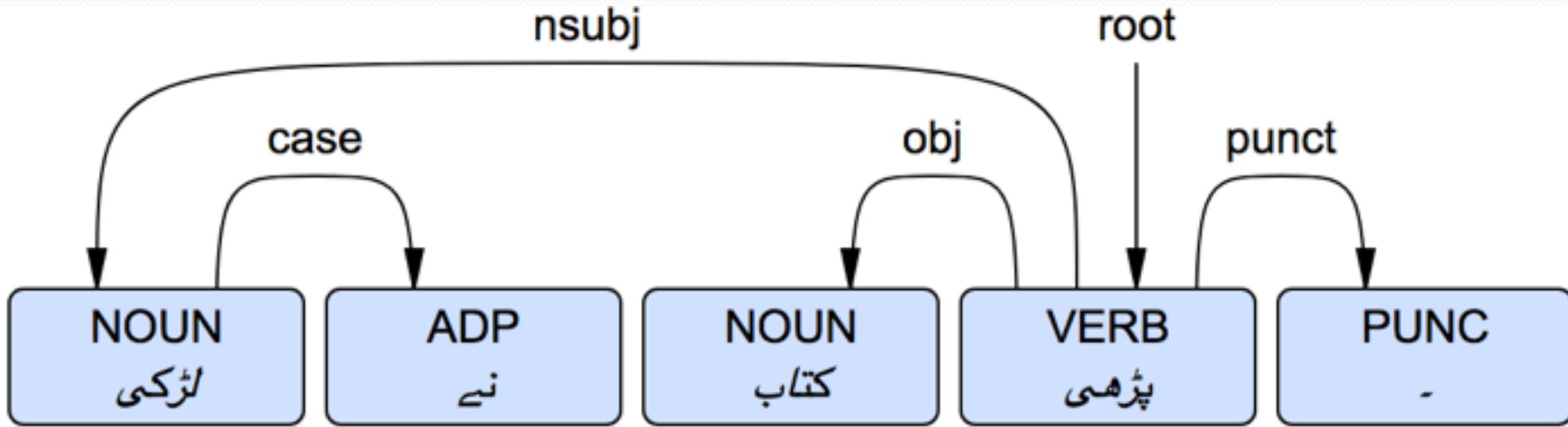
1	لڑکا	لڑکا	NOUN	—	—	3	<u>nsubj</u>
2	کتاب	کتاب	NOUN	—	—	3	obj
3	پڑھ	پڑھتا	VERB	—	—	0	root
4	ہے	ہے	AUX	—	—	3	aux
5	.	.	PUNC	—	—	3	<u>punct</u>

لڑکا کتاب پڑھتا ہے ۔

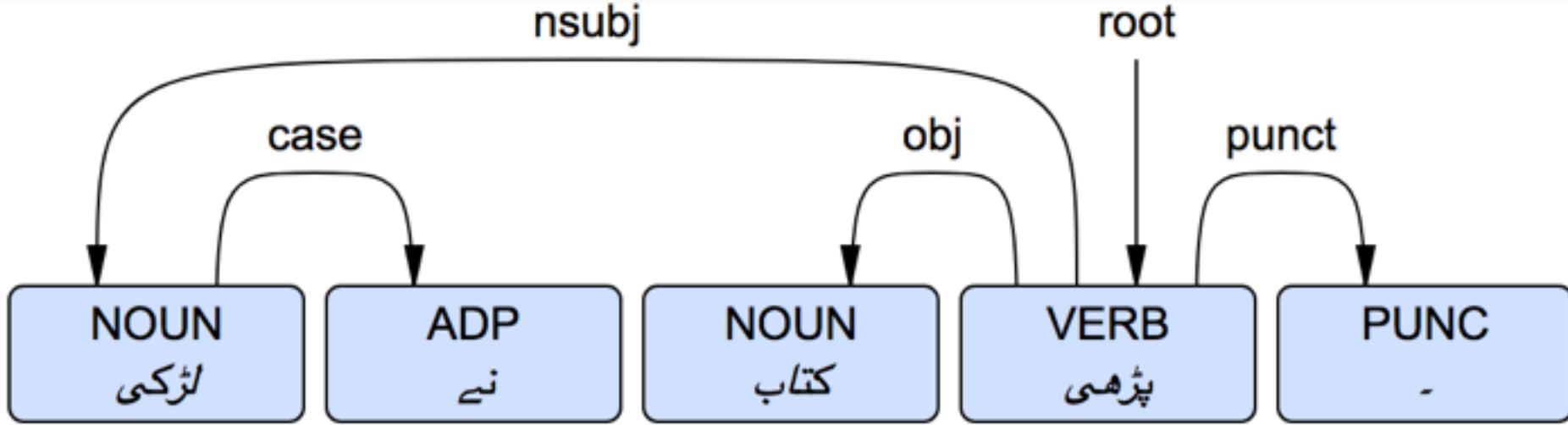
Viewer: <https://urd2.let.rug.nl/~kleiweg/conllu/>



لڑکی نے کتاب پڑھی۔



لڑکی نے کتاب پڑھی۔



ID	FORM	LEMMA	UPOS	XPOS	FEAT	HEAD	DEPREL	DEP	MISC
1	لڑکی	لڑکی	NOUN	—	—	4	<u>nsubj</u>	—	—
2	نے	نے	ADP	—	—	1	case	—	—
3	کتاب	کتاب	NOUN	—	—	4	obj	—	—
4	پڑھی	پڑھ	VERB	—	—	0	root	—	—
5	۔	۔	PUNC	—	—	4	<u>punct</u>	—	—

لڑکے نے سبق یاد کیا۔
لڑکی نے میز صاف کی۔

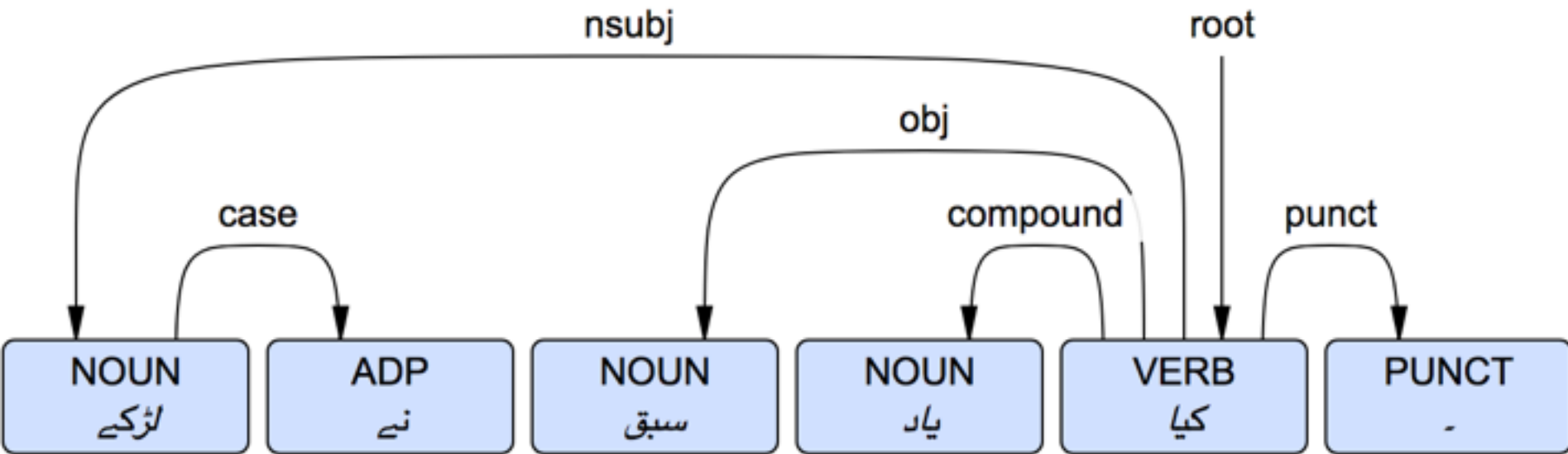
N/Adj+V Complex Predicates

لڑکے نے سبق یاد کیا۔
لڑکی نے میز صاف کی۔

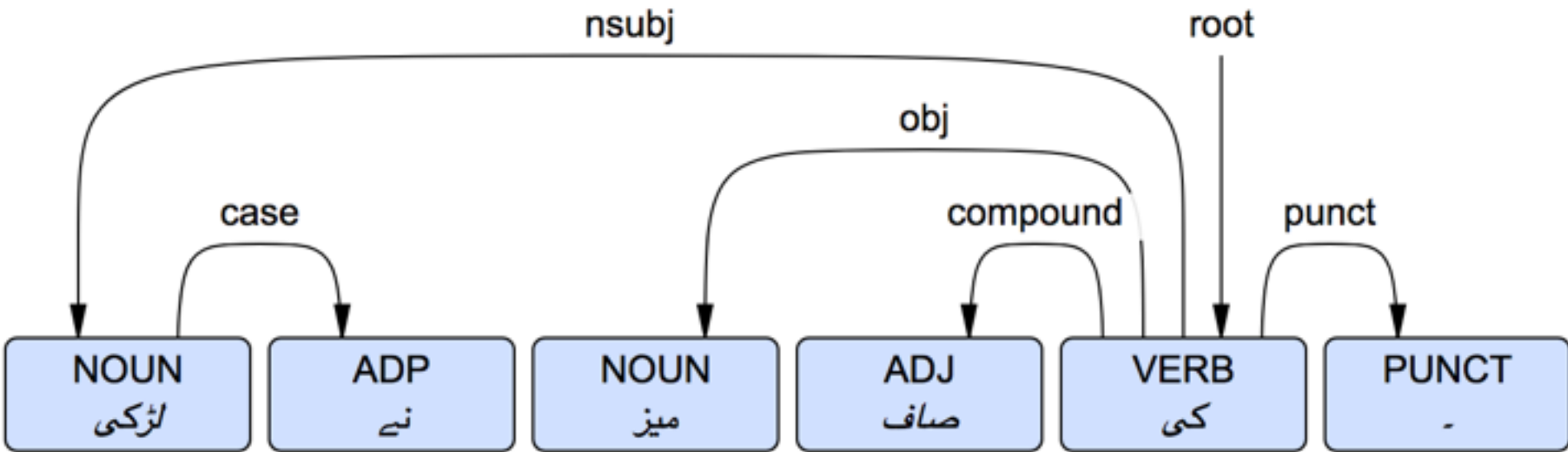
A Reference Dependency Bank for Analyzing Complex Predicates

https://kops.uni-konstanz.de/bitstream/handle/123456789/23101/Ahmed_231017.pdf

لڑکے نے سبق یاد کیا۔



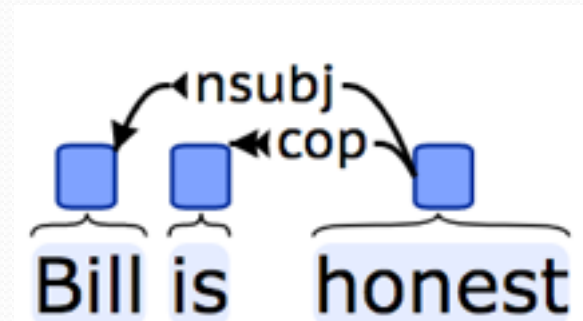
لڑکی نے میز صاف کی۔



لڑکی ذہین ہے۔
لڑکا گھر پر ہے۔

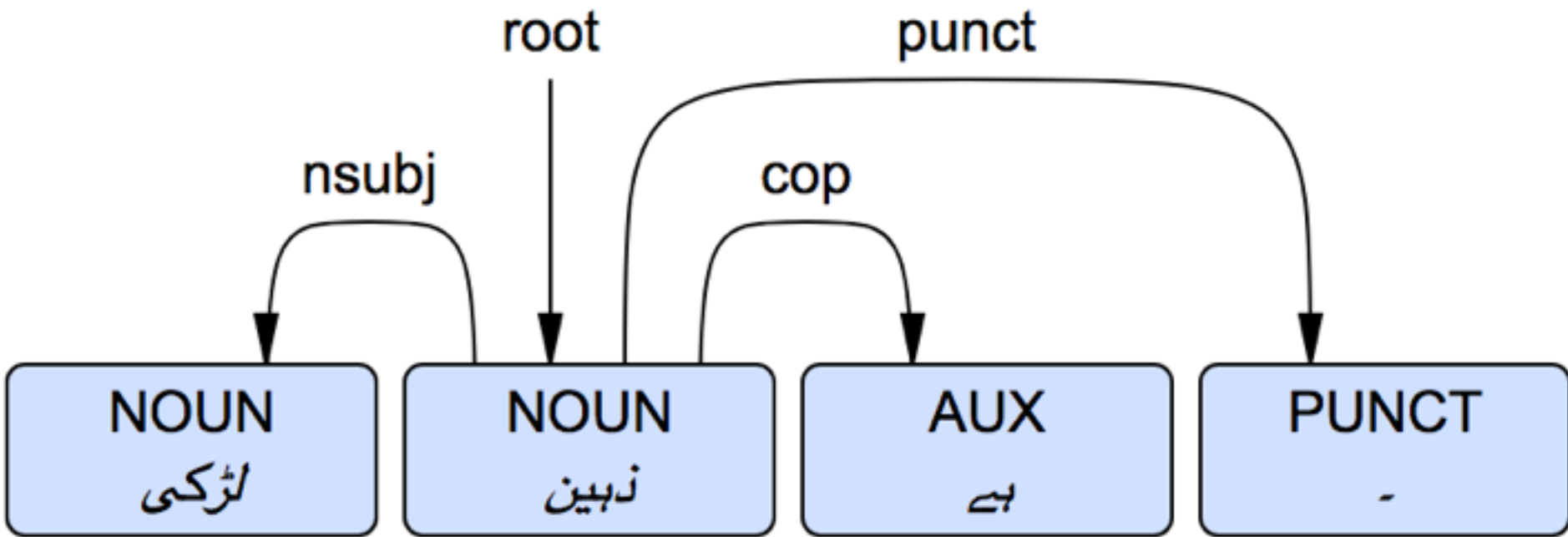
Copula(r Verbs)

- A cop (copula) is the relation of a function word used to link a subject to a nonverbal predicate.
- It is often a verb.

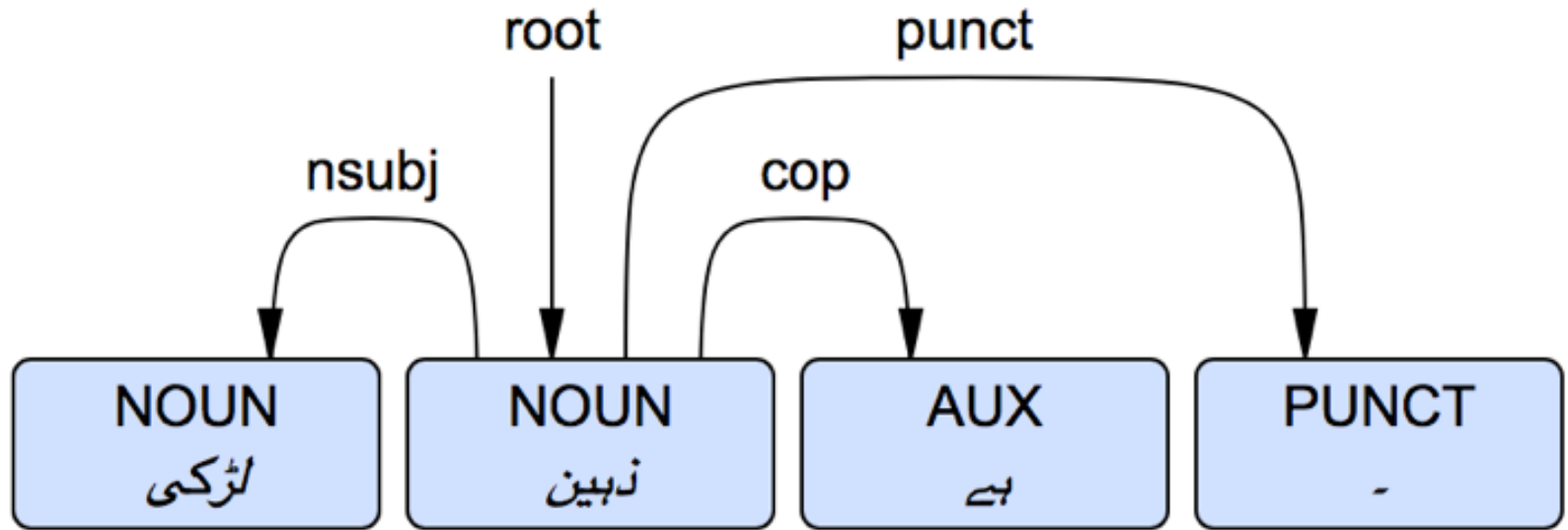


- Such an analysis is motivated by the fact that many languages (e.g. Arabic) often or always lack an overt copula in such constructions.

text = لڑکی ذہین ہے -

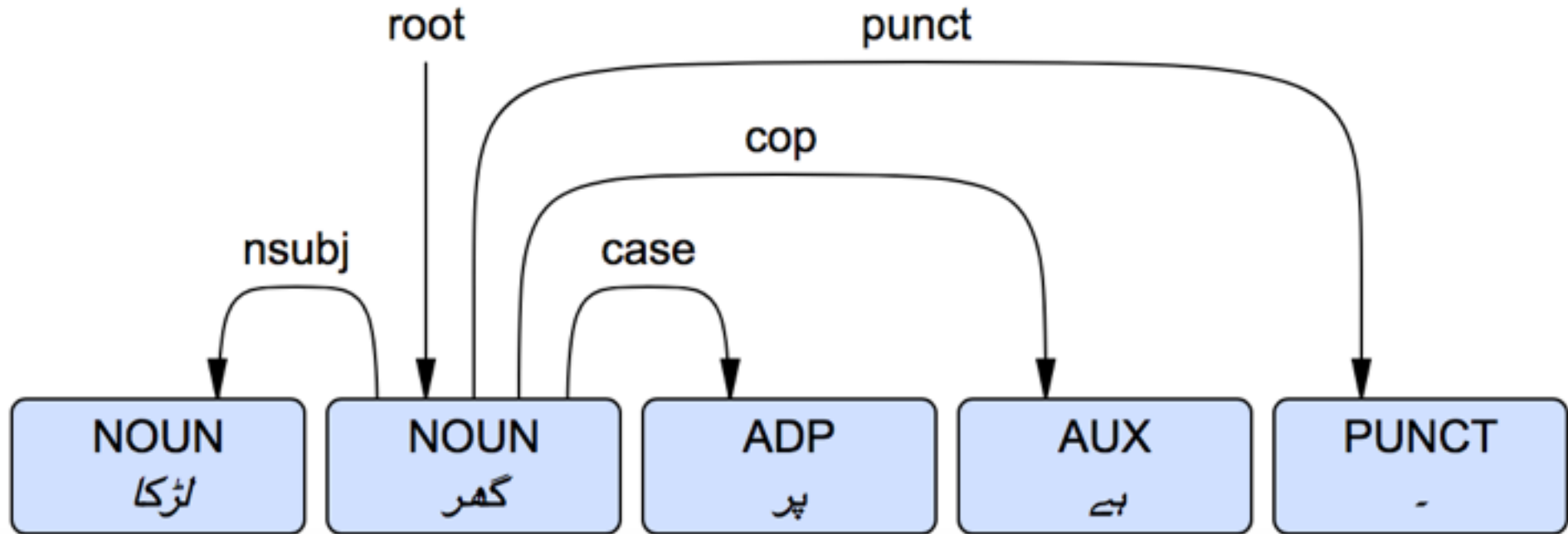


text = لڑکی ذہین ہے۔

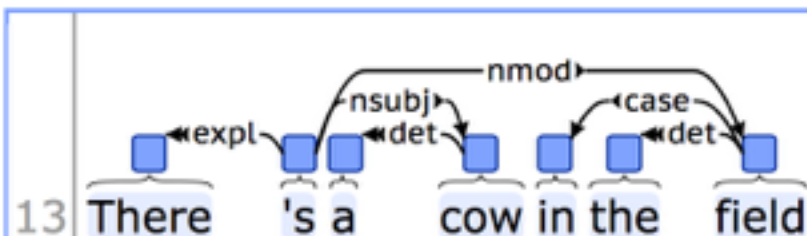
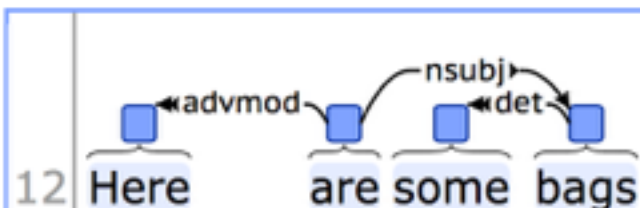
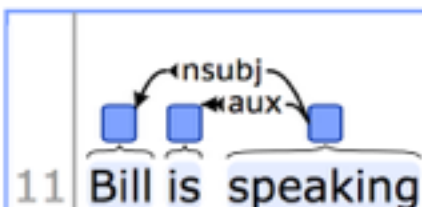
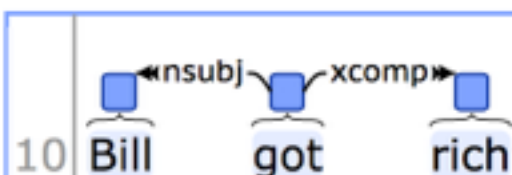


ID	FORM	LEMMA	UPOS	XPOS	FEAT	HEAD	DEPREL	DEP	MISC
# text =	لڑکی ذہین ہے۔								
1	لڑکی	لڑکی	NOUN	—	—	2	<u>nsubj</u>	—	—
2	ذہین	ذہین	NOUN	—	—	0	root	—	—
3	ہے	ہے	AUX	—	—	2	cop	—	—
4	.	.	PUNCT	—	—	2	<u>punct</u>	—	—

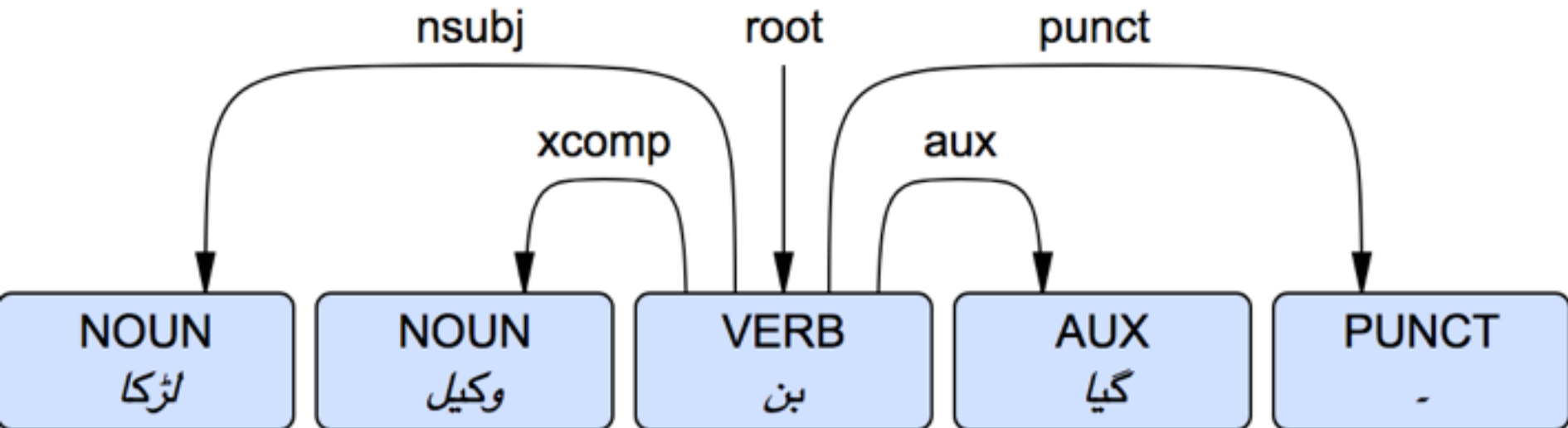
text = لڑکا گھر پر ہے ۔



Predicative “be” is the only verb recognized as a copula; other copula-like verbs, such as “become”, “get”, and “seem”, are treated as regular raising verbs, and thus take xcomp arguments. Non-predicative uses of “be”—e.g., “be” when used in periphrastic verbal constructions, presentationals, or existentials—is annotated as an aux instead of a cop.



text = لڑکا وکیل بن گیا۔



UD - recap

	Nominals	Clauses	Modifier words	Function Words
Core arguments	<u>nsubj</u> <u>obj</u> <u>iobj</u>	<u>csubj</u> <u>ccomp</u> <u>xcomp</u>		
Non-core dependents	<u>obl</u> <u>vocative</u> <u>expl</u> <u>dislocated</u>	<u>advcl</u>	<u>advmod</u> * <u>discourse</u>	<u>aux</u> <u>cop</u> <u>mark</u>
Nominal dependents	<u>nmod</u> <u>appos</u> <u>nummod</u>	<u>acl</u>	<u>amod</u>	<u>det</u> <u>clf</u> <u>case</u>
Coordination	MWE	Loose	Special	Other
<u>conj</u> <u>cc</u>	<u>fixed</u> <u>flat</u> <u>compound</u>	<u>list</u> <u>parataxis</u>	<u>orphan</u> <u>goeswith</u> <u>reparandum</u>	<u>punct</u> <u>root</u> <u>dep</u>