



## Urdu Parts of Speech (POS) Tagset

Parts of Speech (POS) tagging is a fundamental component of most natural language processing systems. The development of Tagset of a language is the first step towards the achievement of this task. The current Urdu Tagset consists of syntactic categories, and improves upon the earlier versions available (Muaz et al. (2009)<sup>1</sup>, Sajjad (2007)<sup>2</sup>, Sajjad and Schmid (2009)<sup>3</sup>). This POS Tagset has been used to develop the CLE Urdu Digest POS Tagged Corpus 100K (available at: <http://cle.org.pk/clestore/urdu Digest corpus 100k tagged.htm>) developed by Center for Language Engineering, KICS, UET, Lahore.

This work has been developed through the project grant for Essential Urdu Linguistic Resources ([www.cle.org.pk/eulr](http://www.cle.org.pk/eulr)) in collaboration with University of Konstanz (<http://www.uni-konstanz.de/>), Germany and funded by German Academic Exchange Service, DAAD (<https://www.daad.org/>), Germany.

---

<sup>1</sup> Muaz, A., Ali, A., Hussain, S. "Analysis and Development of Urdu POS Tagged Corpora", in the Proceedings of the 7th Workshop on Asian Language Resources, IJCNLP'09, Suntec City, Singapore, 2009

<sup>2</sup> Sajjad, H. **Statistical Part of Speech Tagger for Urdu**, National University of Computer and Emerging Sciences, Lahore, Pakistan, 2007

<sup>3</sup> Sajjad, H.; Schmid, H. **Tagging Urdu Text with Parts of Speech: A Tagger Comparison**, in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09)*, 2009

## Urdu POS Tagset

This section presents the Urdu POS Tagset. This Tagset includes a total of twelve POS categories divided into thirty-five tags. Specific examples for each type of POS category are provided in the next section below.

Categories	Types	POS Tag
1. Noun	1.1 Common	NN
	1.2 Proper	NNP
2. Verb	2.1 Main Verb Infinitive	VBI
	2.2 Main Verb Finite	VBF
3. Auxiliary	3.1 Aspectual	AUXA
	3.2 Progressive	AUXP
	3.3 Tense	AUXT
	3.4 Modals	AUXM
4. Pronoun	4.1 Personal	PRP
	4.2 Demonstrative	PDM
	4.3 Possessive	PRS
	4.4 Relative Demonstrative	PRD
	4.5 Relative Personal	PRR
	4.6 Reflexive	PRF
	4.7 Reflexive Apna	APNA
5. Nominal Modifier	5.1 Adjective	JJ
	5.2 Quantifier	Q
	5.3 Cardinal	CD
	5.4 Ordinal	OD
	5.5 Fraction	FR
	5.6 Multiplicative	QM
6. Adverb	6.1 Common	RB
	6.2 Negation	NEG
7. Adposition	7.1: Preposition	PRE
	7.2: Postposition	PSP
8. Conjunction	8.1 Coordinate Conjunction	CC
	8.2 Subordinate Conjunction	SC
	8.3 SCKar	SCK
	8.4 Pre-sentential	SCP
9. Interjection	9.1 Interjection	INJ
10. Particle	10.1 Common	PRT
	10.2 Vala	VALA
11. Symbol	11.1 Common	SYM
	11.2 Punctuation	PU
12. Residual	12.1: Foreign Fragment	FF

## Examples and Usage

This section presents examples for each type of POS category defined in the Urdu POS tagset presented in the section above. The example sentences have been extracted from the POS tagged version of the CLE - Urdu Digest Corpus 100k presented in Urooj et al. (2012)<sup>4</sup> and available at: <http://www.cle.org.pk/Publication/papers/2012/7.pdf>

### 1.1 Common Noun

یہ / PDM < لڑکا / NN > اچھا / JJ ہے / VBF لڑکا، بوتل، دیوار	
گردن / NN گھما / VBF کر / SCK < باورچی خانے / NN > کی / PSP سمت / NN نگاہ / NN اٹھانی / VBF	باورچی خانہ، چابی وابی
اگر / SCP < غیر ملکی / NN > ملوث / NN ہیں / AUXT تو / INT یہ / PRP بھی / INT ہماری / PRS کوتاہیوں / NN کا / PSP نتیجہ / NN ہے / VBF	نوجوان، غیر ملکی، جابلوں
سمندر / NN کی / PSP سطح / NN سے / PSP صرف / JJ چند / Q میٹر / NN < اوپر / NN > اُبھر / VBF آیا / AUXA تھا / AUXT	آگے، پیچھے، اوپر، نیچے، سامنے
مصریوں / NN نے / NN سب / NN سے / PSP < پہلے / NN > باہر / NN کو / PSP بطور / RB علاج / NN استعمال / NN کیا / VBF	پہلے، اب، جب، تب، کب، کل
عربی / NNP زبان / NN کی / PSP اس / PRP شدت / NN کے / PSP باعث / NN < وہاں / NN > ایک / CD لطیف / NN بھی / INT ہمارے / PRS ساتھ / RB ہو / VBF گیا / AUXA	یہاں، وہاں، ادھر، اُدھر
قومی / JJ لیدر / NN کے / PSP < قریب / NN > ہی / PRP پارلیمنٹ / NN کی / PSP شاندار / JJ عمارتیں / NN ہیں / VBF	باہر، اندر، قریب، دُور، نزدیک
وہ / PRP اٹھنے / VBI کی / PSP کوشش / NN کرتا / VBF تو / SC اس / PRP کی / PSP گردن / NN < دائیں / NN > < بائیں / NN > جھولنے / VBI لگتی / AUXA	دائیں، بائیں

<sup>4</sup> Urooj, S., Hussain, S., Adeeba, F., Jabeen, F. and Parveen, R. "CLE Urdu Digest Corpus", in the Proceedings of Conference on Language and Technology 2012 (CLT12), Lahore, Pakistan. (URL: <http://www.cle.org.pk/clt12>)

## 1.2 Proper Noun

ساجد، پاکستان، چناب	اللہ تعالیٰ / NNP نے / PSP <پاکستان / NNP / کو / PSP دنیا / NN کی / PSP ہر / JJ نعمت / NN / سے / PSP نواز / VBF ہے / AUXA
---------------------	---

## 2.1 Main Verb Infinitive

بھاگنا، بھاگنے، بھاگنی	بھاگ / VBF کر / SCK <بھاگنے / VBI> کار / PSP انداز / NN
------------------------	---

## 2.2 Main Verb Finite

بھاگتا، بھاگتے، بھاگتی، بھاگتیں، بھاگا، بھاگے، بھاگی بھاگیں، بھاگیں، بھاگوں، بھاگ، بھاگو، بھاگیو، بھاگیے	ہم / PRP بھاگنے / VBI سے / PSP <بھاگتے / VBF> میں / AUXT سرداران / NN سندھانوالیہ / NNP شہر / NN کی / PSP جانب / NN <بھاگے / VBF> ساتھی / NN اسے / PRP میدان / NN جنگ / NN میں / PSP چھوڑ / VBF کر / SCK <بھاگ / VBF> گئے / AUXA ہوں / AUXT
---	--

## 3.1 Aspectual Auxiliary

آتا، آتی، آتے، آئیں، آیا، آئے، آئی، آئیں، آؤں، آؤ، آئیو، آئیے	ذہن / NN میں / PSP تصویر کشی / NN کار / PSP انوکھا / JJ مصرف / NN ابھر / VBF <آیا / AUXA> یہاں / NN بھی / PRT ہر / JJ طرف / NN فوارے / NN چلتے / VBF نظر / NN <آتے / AUXA> میں / AUXT
چلتا، چلتی، چلتے، چلتیں، چکا، چکے، چکی، چکیں، چکوں، چک، چکو، چکیو، چکیے	۱۹۳۳ / CD / SYM سے / PSP بھاری / JJ پانی / NN دریافت / NN ہو / VBF <چکا / AUXA> تھا / AUXT وہ / PDM افراد / NN جو / PRR سگریٹ نوشی / NN ترک / NN کر / VBF <چکے / AUXA> تھے / AUXT

## 3.2 Progressive Auxiliary

رہا، رہی، رہے، رہیں، رہوں، رہ، رہو، رہیو، رہیے	مجمہ / NN توڑا / VBF جا / VBF <AUXP / رہا / > تھا / AUXT کیا / RB غضب / NN کر / VBF <AUXP / رہے / > ہو / AUXT
---	--

## 3.3 Tense Auxiliary

ہے، ہو، ہوں، تھا، تھی، تھے، گا، گی، گے	کسی / PRP نے / PSP خوب / JJ کما / VBF <AUXT / ہے / > بوڑھا / JJ شیخ / NN جو / PRR اپنا / APNA بہترین / JJ لباس / NN پہنے / VBF ہوئے / AUXA <AUXT / تھا / >
---	--

## 3.4 Modal Auxiliary

سکنا، سکتا، سکتی، سکتے، پانا، پڑنا، ہونا	بیاہیوں / NN سے / PSP محفوظ / JJ رہ / VBF <AUXM / سکتا / > ہے / AUXT
چاہیے	صحیح / NN میں / PSP بالکل / RB نہیں / NEG سونا / VBI <AUXM / چاہیے / >

## 4.1 Personal Pronoun

میں، مجھ، مجھے، تم، تو، تجھ، تجھے، تھیں، آپ، انہیں، انہوں، انہی، ہم، ہمیں، ہی	<PRP / میں / > ان / PRP کی / PSP نشاندہی / NN کر / VBF دوں / AUXA / گار / AUXT <PRP / آپ / > نے / PSP یہ / PDM جملہ / NN اکثر / Q افراد / NN کو / PSP کہتے / VBF سنار / AUXA ہو / AUXM / گار / AUXT
یہ، وہ، اس، اُس، ان، اُن، اُسے، اِسے	<PRP / یہ / > کس / PRP کی / PSP ملکیت / NN ہے / VBF
کون، کیا، کب، کہاں، کدھر، کس	آخر / RB ایسا / PDM غصہ / <NN / کون / PRP > سی / PRT خوبی / NN ہے / VBF

## 4.2 Demonstrative Pronoun

یہ، وہ، اس، اُس، ان، اُن، ایسا، ویسا، کیسا، اُسے، وہی، یہی، اسی، اسی	<PDM / معرکہ / NN / سر / NN / کر / VBF / کے / SCK / مجھے / PRP / بڑی / JJ / خوشی / NN / ہوئی / VBF /
---	---

## 4.3 Possessive Pronoun

میرا، میری، تیرا، تیری، ہمارا، ہماری، تمہارا، تمہاری، اسکا، اسکی، انکا، انکی	<PRS / غیر معمولی / JJ / دلچسپی / NN / کو / PSP / بھانپ / VBF / گئے / AUXA /
---	--

## 4.4 Relative Demonstrative Pronoun

جس، جو، جن	<PRD / لوگ / NN / سونا / NN / چاندی / NN / جمع / NN / کر / VBF / کے / SCK / رکھتے / VBF / ہیں / AUXT /
------------	---

## 4.5 Relative Personal Pronoun

جو، جیسا، جنوں، جنہیں، جسے، جس، جدھر، جہاں، جب	وہ / PDM / صبح / NN / <جو / PRR / آج / NN / سے / PSP / ڈیڑھ / FR / صدی / NN / قبل / NN / میرٹھ / NNP / کی / PSP / ایک / CD / چھاؤنی / NN / میں / PSP / طلوع / NN / ہوئی / VBF /
---	---

## 4.6 Reflexive Pronoun

خود، اپنے آپ، بذاتِ خود، آپ	<PRF / اٹھ / VBF / کر / SCK / میرے / PRP / پاس / NN / آئے / VBF /
-----------------------------	---

## 4.7: Reflexive Apna Pronoun

اپنا، اپنی، اپنے	ناموشی / NN / سے / PSP / <اپنا / APNA / کام / NN / کر / VBF / رہی / AUXP / میں / AUXT /
------------------	---

## 5.1 Adjective

بلند، بلندتر، بلندترین، برا، روتا	خود / PRF / بھی / PRT / <بلند / JJ / شخصیت / NN / کے / PSP / مالک / NN / تھے / AUXT /
-----------------------------------	---

## 5.2 Quantifier

بہت، کم، اتنا، جتنا، کتنا، چند، کچھ، کئی، کافی، اکثر، بعض	اپنی / APNA / تحریر / NN / میں / PSP / سے / PSP / <چند / Q> / اقتباسات / NN / ملاحظہ / NN / فرمائیے / VBF
--	--

## 5.3 Cardinal

ایک، دو، تین، ۳، ۵	<تین / CD> / مانگے / NN / لے / VBF / کر / SCK / اسٹیشن / NN / پہنچ / NN / بائیں / VBF
--------------------	---

## 5.4 Ordinal

پہلا، پہلی، پہلے، دوسرا، تیسرا، یکم، دوم	اُردو / NNP / ڈائجٹ / NN / کا / PSP / <پہلا / OD> / شماره / NN / نومبر / NNP / ۱۹۶۰ / CD / ۶ / NN / SYM / میں / PSP / شائع / NN / ہوا / VBF / تھا / AUXT
---	---

## 5.5 Fraction

نصف، آدھا، پونہ، چوتھائی، ڈیڑھ، ڈھائی، ساڑھے	میں / PRP / <آدھے / FR> / گھنٹے / NN / میں / PSP / پہنچ / VBF / گیا / AUXA
---	--

## 5.6 Multiplicative

گنا، دگنا، تہرا	دن / NN / <دگنی / QM> / رات / NN / <چوگنی / QM> / ترقی / NN / عطا / NN / فرمائے / VBF
-----------------	---

## 6.1 Common Adverb

تقریباً، آہستہ، کیا، کیوں، یوں، جوں، کیسے، کیونکر، شاید، ضرور، پھر، تو نہیں	<تقریباً / RB> / دور / CD / کروڑ / CD / سیاح / NN / اے / PRP / دیکھنے / VBI / آتے / AUXA / میں / AUXT
--	--

## 6.2 Negation Adverb

نہیں، نہ، مت	زبان / NN / کے / PSP / ہلانے / VBI / میں / PSP / ذرا / JJ / بھی / PRT / تکلیف / NN / <نہیں / NEG> / ہوتی / VBF
--------------	---





## 9.1: Interjection

واہ، کاش، سبحان اللہ، ماشا اللہ، السلام علیکم، ارے، اے، اوئے	<INJ / ! / PU / PDM / تور / INT / اتنا / Q / پرانا / لڑ زمانہ / NN / ہے / VBF /
---	---

## 10.1: Common Particles

بھی، تو، ہی، سہی، سا، نا	مگر / SCP / پھر / NN / فوراً / RB / <ہی / PRT / خیال / NN / آیا / VBF /
--------------------------	---

## 10.2: Vala Particle

والا، والی، والے، والوں، والو	طَب / NN / مشرق / NNP / میں / PSP / جوڑوں / NN / کا / PSP / درد / NN / دُور / NN / کرنے / VBI / <والی / VALA / عمدہ / JJ / دو انیاں / NN / موجود / JJ / میں / VBF /
-------------------------------	--

## 11.1: Common Symbols

ء، \$، %، +، ،	۱۸۹۶ / CD / <ء / SYM / کی / PSP / اولیکس / NNP /
----------------	--

## 11.2: Punctuation Symbols (Marks)

!، :، (، )، ؟، -، ،	سیاحت / NN / کا / PSP / شوق / NN / مجھے / PRP / بچپن / NN / سے / PSP / ہے / VBF / < / PU /
---------------------	--

## 12.1: Foreign Fragment

بسم اللہ الرحمن الرحیم۔ گرتوخواہی مسلمان زیستن۔ ڈبھرتوم۔ علم الجھائش -	جس / PRR / پر / PSP / تحریر / NN / تھا / VBF / ڈبھرتوم۔ علم الجھائش / FF /
--	--