



**Parts-of-speech Tagset for Nepali**  
(Developed at Madan Puraskar Pustakalaya)

Contributors to the version 1.0 :

Balaram Prasain, Central Department of Linguistics, T.U, Kirtipur  
Laxmi Prasad Khatiwada, Bal Krishna Bal, Prajol Shrestha  
Madan Puraskar Pustakalaya, Lalitpur, Patandhoka,

Amendments proposed by:

Ramraj Lohani, Central Department of Linguistics, T.U, Kirtipur  
Laxmi Prasad Khatiwada, Bal Krishna Bal, Madan Puraskar Pustakalaya,

## **Background**

Nepali is a morphologically rich language. A morphological analyzer analyzes the morphemes present in a word by breaking down the corresponding lexical items and annotating the grammatical categories to these items. By grammatical categories, we generally understand the parts of speech (POS) categories or lexical categories - Noun, Verb, Pronoun, Adverb, Preposition or Postposition, Adjective, Classifier etc. These parts of speech categories have indeed a vital role in natural language analysis and processing as they are greatly useful in determining the likelihood of the occurrence of word(s) in a particular vicinity of any given word. Such prediction is very important in developing several NLP applications like the Grammar Checker, Machine Translation etc (Jurafsky and Martin, 2000). From this perspective, the development of a POS Tagset ( a set of well defined POS Tags for a language) is very essential. The POS Tagset can serve multiple purposes, in manual POS Tagging words or automated POS Tagging system as reference material for both cases.

In the following sections, we will be discussing the development of the POS Tagset for Nepali. One of the crucial issues that needs to be subtly addressed while designing a POS Tagset is its size. Generally, the assumption is - the smaller the tagset, the greater the accuracy. However, in saying so, we would not want to miss the compulsory categories evident in the language and at the same time also not necessarily increase the size of the tagset whenever economy can be maintained. Hence, a middle ground has been adopted while designing a POS Tagset for Nepali. The Tagset for Nepali currently includes 43 tags and covers almost all the grammatical categories in the Nepali language. While developing the Tagset, the Penn Treebank Tagset has been consulted. Much exercise has been done to conform with the tag namings of the Penn Treebank Tagset for the Nepali Tagset as well wherever applicable and practicable. Among other issues, the first priority has been given to the syntactic aspect of the Nepali language rather than the semantic aspect while devising the tags. It has been assumed that the semantic aspects inherent in a particular word would be handled by the creation of relevant rules and other methods as required by the particular application. We looked for the tokenizability of the text i.e., we did not separate the tokens which are morphologically realized, for example, negation and emphasis, which have not been provided separate tags right now. Similarly, a word may be assigned a different tag depending upon context. Hence, while lexical categorization continues to remain our primary tagging method, the syntactic meaning carried by the word contextually will also be taken into consideration.

The grammatical categories Person, Number and Gender are found distinctively in pronouns and adjectival and verbal inflections. In this tagset, we, however, have not considered these distinctions. That is, both masculine and feminine pronouns get the same tag. Similarly, adjectives inflected for masculine and feminine references receive the same tag.

Other remaining issues that are considered while designing this tagset have been discussed in the relevant sections.

Language: Nepali

---

ID No: 1

Part of speech: Common Noun

Tag: NN

Category: Noun

Common nouns refer to person, place or thing in general. NN tag is assigned to denote common nouns.

Example:

Sentence in Nepali: /NN

Analyzed Word constituents: Hari-ERG brother book  
gave.

Translation in English: Hari gave his brother a book.

ID No: 2

Part of Speech: Proper Nouns

Tag: NNP

Category: Noun

Proper nouns are used for referring to specific person, historical documents, institutions, religions and their holy texts etc. NNP tag is given for proper noun.

Example:

Sentence in Nepali: /NNP /NNP  
/NN

Analyzed Word constituents : Hari-ERG Sarita book  
gave.

Translation in English: 'Hari gave Sarita a book.'

## 2. Pronoun

Besides, person, number and gender, pronouns also distinguish honorificity, which too we are not considering in this tagset. We have three categories of pronouns - personal, possessive and reflexive. First person and second person pronouns are handled within the Personal



reflexive( ) pronouns. Possessive reflexive pronouns inflect for gender and number which we are not distinguishing here.

Example:

Sentence in Nepali: / PPR

Analyzed word constituents: I-REFL PART rice be-eaten.

Translation in English: 'Rice is eaten by me.'

### 3. Demonstratives

In Nepali, some forms function as third person pronouns and demonstratives. For example, in is demonstrative whereas in , it is third person pronoun. Besides, demonstratives also include other forms that inflect for gender and number like , , etc., and that do not inflect for gender and number, i.e., , . So, we have two categories of Demonstratives - marked and unmarked.

ID No: 6

Part of speech: Marked Demonstrative

Tag: DM

Category: Demonstrative

Marked Demonstratives include those which inflect for gender and number.

Example:

Sentence in Nepali: /DM

Analyzed word constituents: Another girl today didn't come

Translation in English: Another girl didn't come today.

Sentence in Nepali: /DM

Analyzed word constituents: Such boy-ERG work don't do

Translation in English: 'Such a boy does not work.'

ID No: 7

Part of speech: Unmarked Demonstrative

Tag: DUM

Category: Demonstrative

Unmarked Demonstratives cover those demonstratives that do not inflect for gender and number. However, they vary in their forms if they are followed by postposition. For example, / 'he' receive the same tag.

Example:

Sentence in Nepali: / DUM

Analyzed word constituents: That my Friend  
is

Translation in English: He is my friend.

Sentence in Nepali: / DUM

Analyzed word constituents: That my Friend  
is.

Translation in English: She is my friend.

#### 4. Verb

Verbal system in Nepali is highly complex in that the verbal conjugation carries gender, number, person and tense, aspect, mood (TAM) information, and compounding. Conjugation also includes markers for passive, causative and negative. For the purpose of our tagset, we have not considered the distinction due to gender, number and person and passivization, causativization and negativization. Regarding compounding, if it appears as a single word then it receives the tag according to the form of the last verb. And if it is due to the composition of separate words, they receive their respective tags as per their forms.

This tagset classifies verbs into three prime categories - finite, nonfinite and auxiliary. Among these, nonfinite ones are further classified into two groups - infinitive and participle. Participles are again divided into three groups - prospective, aspectual and others.

ID No: 8

Part of speech: Finite verb

Tag: VBF

Category: Verb

Finite Verbs include verbs with tense inflections. By implication, it assumes imperatives and optatives as well. VBF is assigned for this group.

Example:

Sentence in Nepali: \_\_\_\_\_ /VBF

Analyzed word constituents in Nepali: Shyam-ERG rice ate

Translation in English: 'Shyam ate rice.'

Sentence in Nepali: \_\_\_\_\_ /VBF

Analyzed word constituents in Nepali: Shyam-ERG rice ate-not

Translation in English: 'Shyam did not eat rice.'

ID No: 9

Part of speech: Auxiliary verb

Tag: VBX

Category: Verb

There are two auxiliary verbs in Nepali - and . These and their corresponding past forms receive separate tag, VBX.

Example:

Sentence in Nepali: \_\_\_\_\_ /VBX

Analyzed word constituents: Ram rice eating was

Translation in English: 'Ram was eating rice.'

ID No: 10

Part of speech: Verb infinitive

Tag: VBI

Category: Verb

Verbs that end with , and that basically refer to the infinitiveness are grouped in this class. VBI is assigned for this class.

Example:

Sentence in Nepali: /VBI

Analyzed word constituents: Hari rice eat home went

Translation in English: 'Hari went home to eat rice.'

Sentence in Nepali: /VBI

Analyzed word constituents: Morning walk beneficial is

Translation in English: 'Walking in the morning is beneficial .

Participle verb

Participles share behavioral properties with other lexical categories such as adjectives and adverbs. Besides, they also come as a member in a verbal chain.

ID No: 11

Part of speech: Prospective participle verb

Tag: VBNE

Category: Verb

This category of verbs that ends in modifies nouns. For example, in " (walking person)", modifies . It also comes to mark habituality in verbal compounding. VBNE is assigned for this group.

Example:

Sentence in Nepali: / VBNE

Analyzed word constituents: Job do man call

Translation in English: Call the working man.

ID No: 12

Part of speech: Aspectual participle verb

Tag: VBKO

Category: Verb

The nonfinite forms of the verbs that end with / are considered as aspectual participles. In Nepali, they may appear as a member in a verbal chain and may play the adjectival role. Though inflects for gender and number, this tagset does not distinguish them. VBKO is given to this group.

Example:

Sentence in Nepali: / VBKO

Analyzed word constituents: You-ERG gave book good  
is

Translation in English: The book that you gave is good.

ID No: 13

Part of speech: Other participle verb

Tag: VBO

Category: Verb

The nonfinite form of verbs that end with / / and / / / are grouped in this class. Basically, these verbs characterize manner.

Example:

Sentence in Nepali: /VBO

Analyzed word constituents: He work finish home  
went

Translation in English: 'He went home after finishing the work.'

Sentence in Nepali: /VBO

Analyzed word constituents: Work do talk do-not

Translation in English: Don't talk while working.

## 5. Adjective

Adjectives, in Nepali, are found in different forms. Some of the adjectives inflect for gender and number, some of them do not inflect, and some inflect for degree. Those that inflect for gender and number have (masculine singular), (feminine singular) and (other, plural) endings. Unmarked adjectives, on the other hand, are neutral in terms of such marking. For example, 'good' can be used for both masculine and feminine. Degree adjectives have comparative and superlative forms. For example, 'more' has comparative and superlative .

So, we have three tags for adjectives in our tagset - JJ for normal adjectives, JJM for marked adjectives and JJD for comparative and superlative forms.

ID No: 14

Part of speech: Normal/unmarked Adjective

Tag: JJ

Category: Adjective

Unmarked adjectives do not inflect for gender and number of the word they modify.

Example:

Sentence in Nepali: /JJ

Analyzed word constituents: Good boy book reading

Translation in English: 'The good boy is reading a book.'

Sentence in Nepali: /JJ

Analyzed word constituents: Good girl Television watching

Translation in English: 'The good girl is watching television.'

ID No: 15

Part of speech: Marked Adjective

Tag: JJM

Category: Adjective

Adjectives ending with , and come in this group.

Example:

Sentence in Nepali:

/JJM

Analyzed word constituents: That boy very nice is  
Translation in English: 'That boy is very nice.'

Sentence in Nepali:

/JJM

Analyzed word constituents: That girl very nice is  
Translation in English: That girl is very nice.

ID No: 16

Part of speech: Degree Adjective

Tag: JJD

Category: Adjective

Comparative forms of adjectives that end with and superlative forms of adjectives that ends with are grouped in this class.

Example:

Sentence in Nepali: /JJD

Analyzed word constituents: highest point

Translation in English: The highest point .

## 6. Adverb

Adverbs are distinguished into two categories. Adverbs that modify verbs are identified as Manner Adverbs in this tagset whereas adverbs that refer to temporal and spatial information are grouped under another class, namely Other Adverbs.

ID No: 17

Part of speech: Manner Adverb

Tag: RBM

Category: Adverb

These adverbs describe the manner of the verbs. RBM tag is assigned to them.

Example:

Sentence in Nepali:

ह्र्दछ

Analyzed word constituents: He slowly walks







Sentence in Nepali: /POP

Analyzed word constituents: Book table above  
is

Translation in English: 'The book is on the table.'

Confusing tags

NN <> POP

Sometimes, postpositions can appear as nouns. In such cases, despite their function, they are assigned the default tag POP. This ambiguity is resolved later by some disambiguation rules. In the first example below, 'downstairs' is functionally noun but this tagset first assigns it a postposition tag POP like in the second example. Later the disambiguation rule retags it as NN.

Sentence in Nepali: /NN

Analyzed Word constituents: Downstairs from dog coming

Translation in English: A dog is coming from downstairs.

Sentence in Nepali: /POP

Word constituents Analysis: Bag table under put.

Translation in English: Put the bag under the table.

VBO <> POP

In Nepali, words like , etc., serve different grammatical categories according to the context. Hence such words initially would get a tag of the default category i.e., VBO which would later be disambiguated accordingly.

Sentence in Nepali: /VBO

Analyzed word constituents: Eat come say he out went

Translation in English: I called him to eat but he went outside.

Sentence in Nepali: - /POP

Analyzed word constituents: Ram than Shyam clever  
is

Translation in English: 'Shyam is more clever than Ram.'

## 9. Conjunction

Conjunctions play a connective role between two or more words, clauses or sentences. There are two types of conjunctions in Nepali - coordinating and subordinating .

ID No: 24

Part of speech: Coordinating Conjunction

Tag: CC

Category: Conjunction

Coordinating Conjunctions is used to join the syntactic elements of the same status. CC tag is assigned for this class.

Example:

Sentence in Nepali: /CC

Analyzed word constituents: I snacks eat and  
office go

Translation in English: I take snacks and go office.

ID No: 25

Part of speech: Subordinate Conjunction

Tag: CS

Category: Conjunction

Subordinating Conjunctions join two or more syntactic elements, one of which is dependent. In Nepali, there are plenty of such subordinating conjunctions. This group consists of elements like , , etc. and relative pronouns like , , which introduce relative clauses. CS tag is assigned to this type of conjunction.

Example:

Sentence in Nepali:

/CS

Analyzed word constituents: He office comes-not because he  
sick is

Translation in English: He doesn't come to office because he is sick.

Example:

Sentence in Nepali:

/CS

Analyzed word constituents: Whoever first comes he-ERG  
rice eats

Translation in English: Whoever comes first will eat rice.

## 10. Interjection

An interjection is a word in a sentence that conveys emotional sense. Interjection will be tagged as UH.

ID No: 26

Part of speech: Interjection

Tag: UH

Category: Interjection

Example:

Sentence in Nepali: !  
!

Analyzed word constituents: Oh! how nice garden

Translation in English: Oh! What a nice garden!

## 11. Number

We have two kinds of numbers - cardinals and ordinals. They are tagged as CD and OD respectively.

ID No: 27

Part of speech: Cardinal Number

Tag: CD

Category: Number

Example:

Sentence in Nepali: /CD

Analyzed word constituents: Five CLAS person  
running

Translation in English: 'Five persons are running.'

ID No: 28

Part of speech: Number ordinal

Tag: OD

Category: Number ordinal

Example:

Sentence in Nepali:

/OD

Analyzed word constituents: Hari exam-in fifth became

Translation in English: Hari became fifth in the exam.

## 12. Plural Marker

This is a single word tag. Although is not a part of speech category in itself, it serves as a plural and collective marker.

ID No: 29

Part of speech: Plural Marker

Tag: HRU

Category: Plural Marker

Example:

Sentence in Nepali:

Analyzed word constituents: Cow came

Translation in English: The cow came.

Sentence in Nepali: - /HRU

Analyzed word constituents: cow-PL came

Translation in English: Cows came.

## 13. Question Word

Question words are used to make questions - both yes/no and content. QW



tag will be used for them.

ID No: 32

Part of speech: Particle

Tag: RP

Category: Particle

Example:

Sentence in Nepali: /RP /RP

Analyzed word constituents: PART I-ERG PART your saying  
understood-not

Translation in English: I didn't understand your saying.

### 15. Determiner

Determiners in Nepali coincide with demonstrative. They are disambiguated according to contexts.

ID No: 33

Part of speech: Determiner

Tag: DT

Category: Determiner

Examples

Sentence in Nepali: / DUM

Analyzed word constituents: That my friend  
is

Translation in English: 'That is my friend.'

Sentence in Nepali: /DT

Analyzed word constituents: That boy my friend  
is

Translation in English: 'That boy is my friend.'

### 16. Unknown Word

Unknown words are those which are not identified in the process of



ID: 36  
Part of speech: None  
Tag: YF  
Category: sentence final ? . !

ID: 37  
Part of speech: None  
Tag: YM  
Category: sentence medial , ; : :- / -

ID: 38  
Part of speech: None  
Tag: YQ  
Category: Quotation ‘ “

ID: 39  
Part of speech: None  
Tag: YB  
Category: Brackets () {} []

## 19. Abbreviation

Abbreviations include those are written with or without dots and other acronyms. FB is assigned for abbreviation.

ID: 40  
Part of speech: Any  
Tag: FB  
Category: Abbreviation

Example:

Example in Nepali: . . ./FB

Example in Nepali: /FB

## 20. Header List

Header List includes all the alphanumeric symbols that are used for

listing purposes. Single characters and characters with or without dots, brackets, comma etc., bullets, stars all get the same tag. ALPH is the assigned tag to this group.

ID: 41  
Part of speech: None  
Tag: ALPH  
Category: Header List

Example:  
) /ALPH , /ALPH , .) /ALPH  
/ALPH, /ALPH, ) /ALPH

## 21. Symbol

Mathematical symbols like +, -, &, % etc receive SYM tag.

ID: 42  
Part of speech: None  
Tag: SYM  
Category: Symbol

Example:  
Example in Nepali: /CD %/SYM  
Translation in English: 2% interest

## 22. Null

This category is devised with the intention of tagging noisy symbols, keywords and control characters which sometimes come along with the text to be POS-Tagged.

ID: 43  
Part of speech: any  
Tag: <Null>  
Category: Null

## Acknowledgment

"This work was carried out with the aid of a grant from the International Development Research Centre, Ottawa, Canada administered through the centre for Research in Urdu Language Processing

(CRLUP), National University of Computing and Emerging Sciences, Lahore, Pakistan (NUCES)”

References:

Daniel Jurafsky & James H. Martin, 2000 A.D : An Introduction to Natural Language Processing. Computational Linguistics and Speech Recognition, Pearson Education, Inc

Hemangaraj Adhikari Dr. 2061 BS : Samasamayika Nepali byakarana, Bidyarthi pustaka Bhandar, Kathmandu, Nepal

Madhav Prasad Pokhrel Dr. 2056 BS : Nepali vakya vigyan, Ekata book Kathmandu, Nepal

Miriam Butt, Traly Holloway Kings, Maria -Eugenia, Nino and Frederique Segnod, 1995 AD : A grammar writers cook book , CLSI publications UK

Narendra Chapagain, 2053 BS : Adhunika Nepali byakarana tatha rachana, Ratna Pustaka Bhandar, Kathmandu Nepal

Yogendra Prasad yadav, Bhim Narayan Regmi, 2059 BS : Bhasha vigyan New Hira books Enterprise, Kathmandu, Nepal