

# CHAPTER 5

---

## Annotation of Manipuri lexical items

This chapter describes the guideline for annotating Manipuri text with part of speech (POS) tags. The POS tag set followed for POS annotation is based on the ILPOST (Indian Language POS Tagset) framework. It is important to remember that the ILPOST is a framework not a tag set by itself. It is a common POS tag set framework that facilitates language specific customization, writing conventions based customization, cross linguistic generalization, reusability across languages as well as application specific customization. The ILPOST framework, hence, allows customization of tags according to the language and/or application specific morphosyntactic information results a tag set based on it.

### 5.1 Previous works in Tagset

UPenn, Brown and C5 are the tagsets for English designed in early 1970s which were mostly simple lists of tags corresponding to the morphosyntactic features, and varied greatly in terms of granularity [33].

CLAWS2 tagset [68] marked an important change in structure of POS tagsets from a flat structure with unitary tags to a hierarchical structure that allowed for decomposable tags. This enabled to tag all the lexical items of the language with distinct grammatical behaviour and a systematic approach.

Several POS tagsets have been developed by a number of research groups working on Indian Languages. Among them very few are available publicly e.g. IIT-tagset, AU-KBC Tamil tagset etc. These tagsets are motivated by specific

research agenda, they differ considerably in terms of morphosyntactic categories and features, tag definitions, level of granularity, annotation guidelines etc [6].

D.S. Thoudam and S. Bandyopadhyay [71] developed a morphology driven Manipuri POS tagger using a flat tagset consisting of 13 tags. Linguistic Data Consortium for Indian Languages (LDC-IL) cell of Central Institute of Indian Languages (CIIL), Mysore has developed POS tagset for fifteen Indian scheduled languages including Manipuri [40].

A hierarchical tagset with language specific attribute values is required to tag the lexical items of a language having large number of affixes, monosyllabic, agglutinative type of language like Manipuri.

## **5.2 Morphosyntactic categories in Manipuri**

Morphosyntactic is a term in linguistics used to refer the grammatical categories or properties for whose definition criteria of morphology and syntax both apply, as in describing the characteristics of words [12]. It can also be defined as the word class and its grammatical constituents of a word.

It is very difficult to determine word classes in Manipuri as we found in English, Hindi, Russian and Tamil etc. except time-stable nouns like চীঙ/cheeng/mountain, ঐশিং/eesing/water, নুমিৎ/numit/sun etc. The determination of word classes in Manipuri is mainly employed by the corresponding affixes. Some bound roots can be formed noun, verb, adjective and adverb by the morphological process like affixation. The details of the word formation process in Manipuri are discussed in section 6.2 of chapter 6.

In Manipuri, the major word class can be categorized as Noun, Pronoun,

Verb, Modifier, Specifier, Demonstrative, Adverb, Participle, Particle, Residual, Punctuation, Numeral and Reduplication.

**(a) Noun (N):** Different Types of Noun and their attributes are given in the following table:

**Table 5.1: Different Types of Noun and their attributes**

Type	Attributes
Common Noun (NC)	Number, Gender, Dimension, Similaritive, Definiteness, Emphatic, Allative, Case Marker, Set (inclusive/exclusive), Copula
Proper Noun (NP)	Number, Dimension, Similaritive, Definiteness, Emphatic, Case Marker, Set (inclusive/exclusive), Copula
Verbal Noun (NV)	Directional, Honorific, Causative, Reflexive, Commutative, Reciprocal, Purposive, Evidential, Certainty, Mood, Modality, Negative, Prospective, Habitual, Inceptive, Aspect, Nominalizer, Number, Distance, Definiteness, Emphatic, allative, Case Marker, Set (inclusive / exclusive), Copula
Spatio-temporal Noun (NST)	Number, Dimension, Similaritive, Emphatic, Allative, Case Marker, Definiteness, Set (inclusive/exclusive), Copula

**(i) Common Noun (NC):** It refers to the words indicating general things, places, ideas and animals etc. Examples of Common Noun along with their morphosyntactic attributes are given below:

উচেকশিংদা → উচেক\_NC শিং\_PL দা\_DAT

uchek-sing-daa / bird-PL-DAT → to birds

নুপাদুদং → নুপা\_NC.MAS দু\_DST দং\_EXL

nupa-du-dang / man-DST-EXL → only that man

*Attributes:* Number, Gender, Dimension, Similaritive, Definiteness, Emphatic, Allative, Case Marker, Set (inclusive/exclusive, etc.)

**Number:** The morphosyntactic markers for plurality in Manipuri are -শিং/-sing, -খোয় / -khoy etc. However, there is no verbal agreement that these suffixes do not entail any inflection in the verbal domain. The suffix -শিং/-sing is added to animate and inanimate objects while -খোয়/-khoy added to personal pronouns and proper nouns to make plural. The following are the examples in which the plural attribute is present:

e.g.; য়ুমশিং → য়ুম\_NC শিং\_PL

yum- sing / house-PL → houses

চীংশিং → চীং\_NC শিং\_PL

cheeng- sing / hill-PL → hills

ঐখোয় → ঐ\_ PPN.1.SG খোয়\_PL

eikhoy/ I-PL → our

উচেক পাইরি / uchek pairi/ bird is flying

উচেকশিং পাইরি / ucheksing pairi/ birds are flying

মহাক চংলি / mahaak chatli / he is going

মখোয় চংলি/ makhoy chatli / they are going

**Gender:** The morphosyntactic markers -অ/-a or -আ / -aa and -ই/-i or -ঈ /-ee are used for indicating male (MAS) and female (FEM). This is also as same as the category ‘number’ in terms of feature agreement, i.e., there is no feature agreement between the noun (controller) and the verb (target) in the sentence (domain). But, the implication of biological gender is determined through the morphosyntactic realization of the above suffixes as shown below:

e.g.; নুপাা\_NC.MAS

nupaa / man

নুপী\_NC.FEM

nupee / woman

নুপা চংলি / nupaa chatli/ man is going

নুপী চংলি / nupee chatli/ woman is going

**Dimension:** The dimension information in Manipuri is reflected in the nominal paradigm through the affixal nature of the demonstratives, proximal (PRX) -সি/-si; and distal (DST) -দু/-du or -তু /-tu.

e.g.; লমপাকসি → লমপাক\_NC সি\_PRX

lampaak-si / field – PRX → this field

য়ুমদু → যুম\_NC দু\_DST

yum-du / house – DST → that house

উচেতু → উচেক\_NC তু\_DST

uchek-tu / bird-DST → that bird

**Similaritive (SIM):** This is the feature which gives the comparative value between the entities of same categories. The marker -কুম/-kum or -গুম/-gum, -কুমা/-kumbaa or -কুম্ব/-kumba or -গুমা/-gumbaa or -গুম্ব/-gumba is used for indicating ‘Similaritive’ value in Manipuri. The examples of Similaritive in Manipuri are given below:

e.g.; মহাক্কুম → মহাক\_PPN কুম\_SIM

mahaak-kum / he or she -SIM → like him or her

নুপাদুগুমা → নুপা\_NC দু\_DST গুম\_SIM না\_NOM

nupaa-du-gum-naa/ man-DST-SIM-NOM → like that man

মহাক্কুমা → মহাক\_PPN কুমা\_SIM

Mahaak-kumbaa / he or she – SIM → like him or her

অসিগুমা → অসি\_PDM গুমা\_SIM

asi-gumba → this- SIM → like this

**Definiteness (DEF):** In Manipuri, the feature of definiteness is marked with the definite marker -দি/-di or -তি / -ti as shown below:

e.g.; চুদি থুম্মী → চু\_NC দি\_DEF থুম\_VR মী\_DEC  
chu-di thum-mee/sugercane- DEF sweet-DEC → sugercane is sweet  
মহাক্তি → মহাক\_PPN তি\_DEF ডাক্তর\_NC নি\_COP  
mahaak-ti doctor-ni / he-DEF doctor-COP → he is a doctor

**Emphatic (EMP):** Emphatic marker in Manipuri is -মক/-mak, which puts stress on the reference of the noun.

e.g.; নুপীদুমক → নুপী\_NC দু\_DST মক\_EMP  
nupi-du-mak / woman – DST – EMP → exactly that woman

**Allative (ALL):** This feature gives the reading of ‘towardness’ from an item. The suffix markers –রোম/-rom or লোম/-lom or সোম/-som are allatives in Manipuri.

e.g.; লম্বীরোমদা → লম্বী\_NC রোম\_ALL দা\_LOC  
lambee – rom-da / road-ALL-LOC → towards the road  
থোঙ্গাল্লোমদা → থোঙ্গাল\_NC লোম\_ALL দা\_LOC  
thongaal-lom-daa / gate-ALL-LOC → towards the gate  
তাংবলসোমদা → ফম্বাক\_NC সোম\_ALL দা\_LOC  
taangbal-som-daa / chair-ALL-LOC → towards the chair

**Case Marker:** In Manipuri case markers are mainly the suffixes indicating Nominative/Ergative -ন/na; Accusative -পু/-pu or -বু /-bu; Genitive -গী/-gee or -কি/-ki or -গি /-gi; Sociative -ক/-ka or -গ/-ga; and Locative -ত/-ta or -তা/-taa or -দ/-da or -দা /-daa; Dative -ঙোন্দ /- ngonda, -ত/-ta, -তা/-taa, -দ/ -

da, -দা/-daa respectively. The compound form -দগী/-dagee or -তগী/-tagee is also found used for indicating Ablative case expressing separation, expulsion, source and direction of movement from one to another.

**Table 5.2: Case Marker and its morphological forms**

Case Marker	Morphological forms
Ergative	-ন/-na
Nominative	-ন, -না / -na, -naa
Accusative	-পু, -বু / -pu, -bu
Genitive	-গী, -কি, -গি / -gee, -ki, -gi
Sociative	-ক, -গ / -ka, -ga
Dative	-ঙোন্দা/ -ngonda, -ত/-ta, -তা/-taa, -দ/-da, -দা/-daa
Instrumental	-ন / -na
Nominalizer	ব, -বা / -ba, -baa
Locative	-দ , -দা, -ত, -তা / -da, -daa, -ta, -taa
Ablative	-দগী / -dagee, -তগী/-tagee

e.g.; রামনা → রাম\_NP না\_NOM  
 raam-naa / Ram-NOM → by Ram  
 যুমদা → যুম\_NC দা\_LOC  
 yum-da / home-LOC → at home  
 অঙাংগী → অঙাং\_NC গী\_GEN  
 angaang-gee / child-GEN → child's

**Set:** This set consists of morphosyntactic types of particles and clitics. They are mainly of the following attributes:

Inclusive (INL) – সু /-su and Exclusive (EXL) -খক/-khak & -তং/-tang

e.g.; অঙাংশিংসু → অঙাং\_NC শিং\_PL সু\_INL  
 angaang- sing-su / child-PL-INL → children also

লৈকাইদুখক → লৈকাই\_NC দু\_DST খক\_EXL

leikaai-du-khak / local – DST-EXL → only that locality

মদুতং → মদু\_PDM তং\_EXL

madu-tang / that-EXL → that only

(ii) **Proper Noun (NP):** Proper nouns are the words denoting the specific names of a person, place, shop, institution, date, day, month, species, etc. A proper noun in Manipuri can be added the following morphosyntactic attributes as suffix:

Number- -খোয় / -khoy (Human Plural Marker), Similaritive- -কুম or -গুম / -kum or -gum, Allative- -লোম/-lom or -রোম / -rom , Definiteness- -দি/-di or -তি /-ti, Emphatic- -মক/-mak, Case markers (as shown in Table 5.2) and Set etc.

e.g.; ইম্ফালদগী → ইম্ফাল\_NP দগী\_ABL

Imphal –dagee / Imphal – ABL → from Imphal

চাউবাখোয় → চাউবা\_NP খোয়\_PL

Chaoba-khoy / Chaoba – PL → Chaoba and others

(iii) **Verbal Noun (NV):** A Verbal noun in Manipuri is derived from the verb root by attaching nominalizer (NMZ) -ব/-ba or -বা-baa or -প/-pa or -পা / paa to the verb root.

e.g.; চাবা → চা\_VR বা\_NMZ

chaa-baa / chaa-NMZ → eating

চৎপা → চৎ\_VR পা\_NMZ

Chat-paa / go-NMZ → going



Here চৎ/chat & তৌ/tou are verb roots and ব/ba & প/pa are moninalizers. The word formation process will be discussed in the chapter 6. The attributes which can be attached to the Verbal Noun are shown in the Table 5.1.

(iv) **Spatio-temporal Noun (NST)**: Nouns that denote space, time and direction etc., are known as Spatio-Temporal nouns (NST). Since they are structurally nominal, they also take same attributes as other types of noun in Manipuri.

e.g.; মথকুদা → মথক\_NST তু\_DST দা\_LOC  
 mathak – tu –da / top – DST-LOC → on top of that  
 মমাঙলোমদা → মমাঙ\_NST লোম\_ALL দা\_LOC  
 mamaang – lom –da / front – ALL – LOC → towards front

(b) **Verb (V)**: The type of Verb and its attributes are shown in the following table

**Table 5.3: Chart of Verb root and its attributes**

Type	Attributes
Verb Root (VR)	Directional, Honorificity, Causative, Reflexive, Commutative, Reciprocal, Evidential, Certainty, Mood, Modality, Negative, Prospective, Habitual, Inceptive, Aspect, Copula, Evaluative

Verbs in Manipuri are derived by affixation of modality, aspect markers etc. to the verb roots.

e.g.; চৎলি → চৎ\_VR লি\_PRG  
 chat – li / go-PRG → going  
 নোকই → নোক\_VR ই\_DCL  
 nok-i → laugh – DCL → laughs

The attributes of Verb root are discussed as follows:

**Directional:** The directional markers which are mainly attached to the verbal roots for indicating the directions of the events or motions of the action performed by the actor are inward –শিন/-sin or –জিন /-jin , outward –থোক/-thok or –দোক / -dok, upward –খত/- khat or –গত /- gat, downward –থ /- tha or –দ / -da

e.g.; থাংদোকলে → থাং\_VR দোক\_OUT লে\_PRF  
thaang-dok-le / lift-OUT-PRF → lifted out  
পুশিনবিয়ু → পু\_VR শিন\_IN বি\_HON য়ু\_IMP  
pu-sin-bi-yu / bring-IN-HON-IMP → please bring inside

**Honorificity (HON):** Irrespective of agreement feature on the verb along with the subject, honorificity in Manipuri is marked through the use of suffix- -পি/-pi or –বি/-bi to give the readings of formal and intimate.

e.g.; চাবিয়ু → চা\_VR বি\_HON য়ু\_IMP  
cha-bi-yu / eat – HON – IMP → please eat  
থকপিয়ু → থক\_VR পি\_HON য়ু\_IMP  
thak – pi – yu / drink – HON – IMP → please drink

**Causative (CAUS):** In Manipuri, the process of causativization gets reflected morphosyntactically in the verbal base. The causative marker is –হন/-han or –হল/- hal

e.g.; তৌহনখি → তৌ\_VR হন\_CAUS খি\_PRF  
tou-han-khi / do-CAUS-PRF → (someone) has made (the job)  
done

কপহল্লি → কপ\_VR হল\_Caus লি\_PRG

kap –hal-li / cry-CAUS-PRG → (somebody) makes (someone)

crying

**Reflexive (REFL):** Reflexive sentences are formed in Manipuri by adding the marker -চ /-cha or -জ/-ja to the verb.

e.g.; লিৎচরি → লিৎ\_VR চ\_REFL রি\_PRG

lit-cha-ri / wear-REFL- PRG → (I am) wearing (myself)

তৌজনি → তৌ\_VR জ\_REFL নি\_COP

tou – ja-ni / do-REFL – COP → (he) will do the job (himself)

**Commutative (COMM):** The commutative construction giving the sense of togetherness is morphologically marked by the suffix -মিন / -min as shown below:

চামিন্নরি → চা\_VR মিন\_COMM ন\_RECI রি\_PRG

chaa-min-na-ri / eat -COMM-RECI-PRG → eating together

**Reciprocal (RECI):** In Manipuri the reciprocal is formed by adding the marker - ন /- na to the verb root as shown below:

য়েংনরি → য়েং\_VR ন\_RECI রি\_PRG

yeng-na-ri / see-RECI-PRG → seeing each other

ফুনরি → ফু\_VR ন\_RECI রি\_PRG

fu-na-ri / beat- RECI-PRG → beating each other

**Negative(NEG):** It is a grammatical element which reverses the truth value of a proposition to which the element is added. In Manipuri, the suffix -ত /-ta or -দ /-da is used as a negative marker.

e.g.; উদরি → উ\_VR দ\_NEG রি\_PRG  
 u-da-ri / see-NEG-PRG → (still) not seen  
 চত্তরে → চৎ\_VR ত\_NEG রে\_PRF  
 chat-ta-re / go-NEG-PRF → has not gone

**Mood:** Mood is the communicative function contained in a sentence or proposition. It expresses the degree or kinds of reality of a proposition, as perceived by the speaker. The speaker performing a speech act utters a sentence to realize his or her communicative intention. Speech act moods in Manipuri are mainly of the following values:

Declarative (DEC) -ই / -i, Suplicative (SUP) -সি/ -si, Imperative (IMP) -উ/-u or -য়ু /-yu, Prohibitive (PROH) -নু / -nu, Permissive (PERM) -সনু / -sanu, Optative (OPT) -কে /-ke or -গে /-ge, Interrogative (INT) -ল/-la, -র/-ra, নো / -no (for wh-questions).

e.g.; য়েংই → য়েং\_VR ই\_DEC  
 yeng-i / see-DEC → see  
 চৎসি → চৎ\_VR সি\_SUP  
 chat-si / go-SUP → lets go  
 চাউ → চা\_VR উ\_IMP  
 cha-u / eat-IMP → do eat  
 ফুনু → ফু\_VR নু\_PROH  
 phu-nu / beat-PROH → don't beat  
 য়েংসনু → য়েং\_VR সনু\_PERM

yeng-sanu / see- PERM → let him/her see

চাণে → চা\_VR গে\_OPT

chaa-ge/ eat-OPT → let me eat

**Evidential (EVI):** That Manipuri overtly marks inferred sources of evidence by the marker –লম/-lam or –রম/-ram as illustrated below:

থল্লম্মী → থক\_VR লম\_EVI মী\_DEC

thak-lam-mee / drink-EVI-DEC → drank

চাৰম্মী → চা\_VR রম\_EVI মী\_DEC

chaa-ram-mee / eat-EVI-DEC → ate

**Modality:** Modality is the speaker's attitude about the event described. One of the grammatical means by which we can express a speaker's attitude towards the status of an event is modality. In Manipuri, modality is expressed by the use of the morpho-obligation (OBL) -তৌ /-tou or -দৌ / -dou and Volition (VOL) -তোয় / -toy or -দোয় / -doy.

e.g.; লাকতৌনি → লাক\_VR তৌ\_OBL নি\_COP

laak-tou-ni / come-OBL-COP → has to come

চৎকদৌনি → চৎ\_VR ক\_POT দৌ\_OBL নি\_COP

chat-ka-dou-ni / go-POT-OBL-COP → has to go

কৎতোয়নি → কৎ\_VR তোয়\_VOL নি\_COP

kat-toy-ni / offer-VOL-COP → has to offer

ইগদোয়নি → ই\_VR গ\_POT দোয়\_VOL নি\_COP

i-ga-doy-ni / write-POT-OBL-COP → has to write

**Aspect:** Aspect is the internal temporal structure of an event expressed by the verb. This indicates whether the event is completed or ongoing etc. In Manipuri, aspectual information is morphologically marked on the verb. The values are progressive (PRG) -রি / -ri or -লি / -li and perfective (PRF) -রে / -re or -লে / -le as shown below:

চাৰি → চা\_VR ৰি\_PRG  
 chaa-ri / eat- PRG → eating  
 থক্লে → থক\_VR লে\_PRF  
 thak- le / drink – PRF → drank

**Inceptive (INC):** It is a distinctive aspectual form expressing the beginning of a state or activity. It mainly expresses the action's orientation towards the limit, i.e., the meaning of limitedness refers not to the action as a whole, but only to its initial phase. In Manipuri, the marker -ৰক or -লক / -rak or -lak is used as an inceptive marker, as shown below:

লাউৰক্লে → লাউ\_VR ৰক\_INC লে\_PRF  
 lau-rak-le → shout-INC-PRF → started shouting  
 কপলক্লে → কপ\_VR লক\_INC লে\_PRF  
 kap-lak-le / cry-INC-PRF → started crying

**Habitual (HAB):** This aspectual category expresses an action which is regularly consistently performed by some entity. In Manipuri, the marker -কন or -গন / -kan or -gan is used as a habitual marker to indicate the action that someone usually or often does.

e.g.; চৎকনলি → চৎ\_VR কন\_HAB লি\_PRG  
 chat-kan-li / go-HAB-PRG → (He) often goes (to market)

তৌগনলি → তৌ\_VR গন\_HAB লি\_PRG  
tou-gan-li / do-HAB-PRG → (We) often do (the work)

**Prospective (PROS):** This is an aspectual form which typically expresses the idea that some event is imminent. In Manipuri, the marker -ল / -la or -র / -ra is the prospective marker primarily indicating the possibility that the event will happen i.e., the prospect of something happening.

e.g.; তৌরগনি → তৌ\_VR র\_PROS গ\_POT নি\_COP  
tou-ra-ga-ni / do-PROS-POT-COP → going to be done  
চলগনি → চং\_VR ল\_PROS গ\_POT নি\_COP  
chat-la-ga-ni / go-PROS-POT-COP → going to be moved

**Copula (COP):** In Manipuri, copula is used in equational sentences where a subject NP is linked to a predicate, copula is found co-occurring with other verbal or nominal categories. The suffix -নি/-ni is a copulative marker in Manipuri.

e.g.; য়েংগনি → য়েং\_VR গ\_POT নি\_COP  
yeng-ga-ni / see-POT-COP → will see  
চংকনি → চং\_VR ক\_POT নি\_COP  
chat-ka-ni / go-POT-COP → will go

It can also be suffixed to Noun, Pronoun and Proper noun.

e.g.; মসি গাড়ীনি → মসি\_PDM গাড়ী\_NC নি\_COP  
masi gaaree-ni/ vehicle-COP → this is a vehicle  
শৈশকপসি মহাকনি → মহাক\_PPR নি\_COP  
seisakpa-si mahaak-ni /singer-PRX he-COP → he is the singer

**Evaluatives:** These verbal elements, denoting a mood-like category, serve to express speaker's attitude towards a statement the truth of which she/he

accepts. In its strict sense, the truth of the proposition is clarified with the use of these evaluative elements. The attitudes so expressed may be surprise, dubitative, confirmation etc. Manipuri exhibits a set of evaluative sentence-final particles which are given below:

Surprise (SURP) -নে/-ne, Dubitative (DUB) -দ/-da, -ত/-ta , -দা/-daa, -তা/-taa, Confirmation (CONF) -এ/-e, -য়ে/-ye, Exasperation (EXAS) -হে/-he, Persuasion (PERSU) -কো/-ko, Complaint (CMPL) -দো/-do , Insistent (INSIS) -বা/-baa

e.g.; তারেনে → তা\_VR রে\_PRF নে\_SURP  
 taa-re-ne / hear-PRF-SURP → heard (surprisingly)  
 চৎকনিদা → চৎ\_VR ক\_POT নি\_COP দা\_DUB  
 Chat-ka-ni-da / go-POT-COP-DUB → will go (doubly)  
 য়েংগনিয়ে → য়েং\_VR গ\_POT নি\_COP য়ে\_CONF  
 yeng-ga-ni-ye / see-POT-COP-CONF → will see (surely)  
 চৎকনিহে → চৎ\_VR ক\_POT নি\_COP হে\_EXAS  
 chat-ka-ni-he / go-POT-COP-EXAS → will go (exasperatedly)  
 চৎলুকো → চৎ\_VR লু\_IMP কো\_PERSU  
 chat-lu-ko / go-IMP-PERSU → go (persuasively)  
 চৎলগনিদো → চৎ\_VR ল\_PROS গ\_POT নি\_COP দো\_CMPL  
 chat-la-ga-ni-do / go-PROS-POT-COP-CMPL → will have to do  
 (erroneously)  
 হৌরসিবা → হৌ\_VR র\_PROS সি\_SUP বা\_INSIS  
 hou-ra-si-baa / begin-PROS-SUP-INSIS → let's start (insistently)

(c) **PRONOUN (P):** A pronoun is a word that can substitute for a noun or noun phrase (or clause) or words of similar types. Pronouns are referential as well as



anaphoric by nature, i.e., they refer back to something previously expressed (linguistic antecedent) either in the sentence or in the discourse. Types and attributes for Pronouns in Manipuri are shown in the following table:

**Table 5.4: Different types of pronoun and its attributes**

Type	Attributes
Personal Pronoun (PPR)	Number, Dimension, Person, Similaritive, Case Marker, Allative, Distributive, Definiteness, Emphatic, Incl/Excl/Allative
Possessive Pronoun (PPS)	Number, Dimension, Person, Similaritive, Case Marker, Allative, Distributive, Definiteness, Emphatic, Incl/Excl/Allative
Demonstrative Pronoun (PDM)	Number, Dimension, Person, Similaritive, Case Marker, Allative, Distributive, Definiteness, Emphatic, Incl/Excl/Allative
Reflexive pronoun (PRF)	Number, Dimension, Person, Similaritive, Case Marker, Allative, Distributive, Definiteness, Emphatic, Incl/Excl/Allative
Reciprocal pronoun (PRC)	Case Marker
Interrogative pronoun (PIN)	Number, Case Marker, Distributive, Definiteness, Emphatic, Incl/Excl

(i) **Personal pronoun (PPR):** Personal pronoun in Manipuri may be singular or plural or dual. The following table shows the different types of Personal pronoun:

**Table 5.5: Personal pronoun and its different types**

	Singular (SG)	Plural (PL)	Dual (DU)
First Person (1)	ঐ/ei, ঐহাক/eihaak	ঐখোয়/eikhoy	ইবানি/ibaani
Second Person (2)	নঙ/nang, নহাক/nahaak	নখোয়/nakhoy	নবানি/nabaani, নখোয়নি/nakhoyni
Third Person (3)	মা/maa, মহাক/mahaak	মখোয়/makhoy	মবানি/mabaani, মখোয়নি/makhoyni

Examples of tagged Personal pronouns in Manipuri are given below:

ঐনি → ঐ\_PPR.1.SG নি\_COP

ei-ni / I-COP → its me

নবানিগীদি → নবানি\_PPR.2.DU গী\_GEN দি\_DEF

nabaani-gee-di / both of you-GEN-DEF → for both of you

মখোয়সিনি → মখোয়\_PPR সি\_PRX নি\_COP

makhoy-si-ni / they-PRX-COP → they are

(ii) **Possessive Pronoun (PPS):** Possessive pronouns are formed by adding a genitive case marker to the personal pronouns e.g. ঐগী/ei-gee/my, ঐহাকী/eihaak-kee/my, ঐখোয়গী/ei-khoy-gee/our, মাগী/maa-gee/his or her, মহাকী/mahaak-kee/his or her, মখোয়গী/makhoy-gee/their, নঙগী/nang-gee/your (Singular), নহাকী/nahaak-kee/your (singular), নখোয়গী/nakhoy-gee/your (plural)

(iii) **Demonstrative Pronoun (PDM):** A demonstrative pronoun is used to refer to a previous noun or an item of similar type in the discourse. Demonstrative Pronouns in Manipuri are অসি/asi, মসি/masi, অদু/adu and মদু/madu.

Examples of tagged demonstrative pronoun in Manipuri are given below:

মদুগীনি → মদু\_PDM.3 গী\_GEN নি\_COP

madu-gee-ni / that-GEN-COP → (this is) for that

অসিৰু → অসি\_PDM.3 বু\_ACC

asi-bu / it-ACC → it is

Personal pronoun, demonstrative pronoun and possessive pronoun are collectively known as pronominal pronoun. They have the same attributes as shown in the Table 5.5. The attributes are discussed as follows:

**Number:** It should be marked as ‘SG’ for singular, ‘PL’ for plural and ‘DU’ for dual if it is morphologically present in the word.

**Person:** Person attribute would be marked in personal pronouns. All the demonstrative pronouns would be marked as ‘3’.

**Case Marker:** Case-Markers should be marked if the morphological attribute is present there.

**Distributive:** This is the feature of which the pronoun occurs twice consecutively in the sentence. This is one of the iterativity features of subjects.

**Definiteness:** For definiteness feature in the pronominal pronouns, it should be marked ‘DEF’ if it is morphologically present in the word.

**Similaritive:** This is the feature which gives the comparative value between the entities of same categories. It should be marked ‘SIM’ if it is morphologically present.

**Allative:** This feature gives the reading of ‘towardness’ from an item. It should be marked ‘ALL’ if it is present in the word.

**(iv) Reflexive Pronoun (PRF):** A reflexive pronoun is a pronoun that is preceded by the noun, adjective, adverb or pronoun to which it refers (its antecedent) within the same clause. It co-refers to the subject or the object in the sentence. Reflexive pronouns in Manipuri are derived by prefixation of pronominal prefixes ই-/i-, ন-/na-, ম-/ma- standing for the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> persons respectively, to the root সা / saa “body” as shown below:

ইসা/i-saa /myself, নসা /na-saa/yourself, মসা/ma-saa/himself or herself

In Manipuri, reflexive pronouns are found in their various nominal inflection patterns, as in following examples:

ইসানা → ইসা\_PRF.1.SG না\_NOM

i-saa-na / PRF.1.SG- body-NOM → by myself

নসামকনা → নসা\_PRF.2.SG মক\_EMPH না\_NOM

na-saa-mak-na / PRF.2.SG -body-EMP-NOM → by (you) yourself

মসাগীনি → মসা\_PRF.3.SG গী\_GEN নি\_COP

ma-saa-gee-ni / PRF.3.SG body-GEN-COP → (it is) for himself

Attributes are shown in the Table 4.4 and all the attributes are discussed in the above.

(v) **Reciprocal Pronoun (PRC):** A reciprocal pronoun is the one which describes a relationship of mutuality between the plural subjects of a sentence. In Manipuri, reciprocal pronoun such as ‘each other’ is expressed by the phrasal form consisting of a numeral ‘one’ plus a nominative case and a numeral ‘one’ plus a locative suffix, as অমনাঅমদা/amanaa-amadaa. Case markers are the attributes of Reciprocal Pronoun.

(vi) **Interrogative Pronoun (PIN):** In Manipuri, interrogative pronoun begins with the prefix ক-/ka- and it is mainly used in asking questions. Number, Case Marker, Distributive, Definiteness, Emphatic, Incl/Excl are attributes associated with interrogative pronoun. Rules for other attributes are as same as other pronouns, i.e., they are marked only if they are morphologically present. Interrogative words of question are given below:

কনা/kana, করি/kari, কয়া/kaya, কদায়/kaday,

কয়াম/kayam, কদোম/kadom, কয়াম/karam etc.

(d) **Modifier (M):** Modifiers are the categories which usually modify nominal constructions in the sentences. In short, this is the process of limiting a

linguistic element by another dependent linguistic element. Adjectives and quantifiers are the nominal modifiers on the basis of their syntactic functions.

The types and attributes of nominal modifiers are given in the following table:

**Table 5.6: Different types of Modifier and its attributes**

Type	Attributes
Adjective (MJ)	Gender, Number, Dimension Allative, Case Marker, Negative, Emphatic, Similaritive, Definiteness, inclusive/exclusive
Quantifier (MQ)	Numeral, Dimension, Similaritive, Case Marker, Definiteness, Emphatic, Inclusive/Exclusive

(i) **Adjective (MJ):** Since an adjective modifies a noun, it is kept as a type of modifier in the framework. In Manipuri, adjectives are morphologically derived from the verbal roots in such a way that an attributive prefix অ- / a- is prefixed to the verbal noun form comprising of a verbal root followed by a nominalizer -ব/-ba or -বী/-baa, as given below:

অ -VR -বা

অফবা / a-pha-baa → good

অশেংবা / a-seng-baa → clean

অঙাংবা / a-ngaang-baa → red

অঙৌবা / a-ngou-baa → white

In Manipuri, an adjective can either precede or follow a noun without altering the sense. In the case of di-syllabic roots, the prefix অ- /a- is not found necessary as ফজবা/phajabaa, চংখোনবা/changkhnabaa, নিংথীবা / ningtheebaa etc.

e.g.; অফবগীনি → অফব\_MJ গী\_GEN নি\_COP

aphaba-gee-ni / good-GEN-COP → for the sake of goodness

ফজবীশিংদুনা → ফজবী\_MJ.FEM শিং\_PL দু\_DST না\_NOM

phajabee-sing-du-naa / beautiful (girl)-PL-DST-NOM → by those  
beautiful girls

(ii) **Quantifier (MQ)**: A quantifier expresses some notion of quantity of the noun in terms of the definite or indefinite number or amount of the noun. Quantifiers in Manipuri follow a noun, which allow them to take some of the nominal inflections.

e.g.; খিত্তং/khittang/little, খরা/khara/few, ময়াম/mayaam/many,

খজিক/khajik/little, য়ান্না/yamnaa/more

The attributes are marked only if they are morphologically visible in the words.

(e) **Specifier (SPEC)**: It specifies a group of nominal and adjectival elements with same characteristics. In Manipuri, Specifier follows a noun or adjective.

e.g. অঙৌবা মথৈ → অঙৌবা\_MJ মথৈ\_SPEC

angoubaa makhei / white-SPEC → all the white (colours only)

অশাংবা মথৈ → অশাংবা\_MJ মথৈ\_SPEC

asaangbaa makhei / long-SPEC → all the long (sticks only)

Some Case Markers are attributes of Specifier.

e.g. মথৈগী → মথৈ\_SPEC গী\_GEN

makhei-gee

মথৈনি → মথৈ\_SPEC নি\_COP

makhei-ni

মথৈদি → মথৈ\_SPEC দি\_DEF

makhei-di

(f) **Demonstratives (D):** Demonstratives form a class of words which deictically refers to the entity or the object. It refers to a spatial deixis for the location of the referent (i.e., object or the entity). Demonstratives are the words that indicate the spatial, temporal or discourse location of its referent i.e., the noun which follows the demonstrative has a definite reference in the discourse. In Manipuri demonstratives follow the nouns.

Types and attribute of demonstratives:

**Table 5.7: Different types of Demonstrative and its attributes**

Types	Attributes
Absolute Demonstrative (DAB)	Case Marker, Dimension, Similaritive, Emphatic, Definiteness, Inclusive/Exclusive
Wh-demonstrative (DWH)	Case Marker, Dimension, Emphatic.

(i) **Absolute Demonstrative (DAB):** Demonstratives have the same form of the pronoun, but they are different in their distribution from the pronouns. In Manipuri, absolute demonstratives follow their nouns.

e.g. নুপা অদুমকনা\_DAB

nupaa adumaknaa / man-DAB → by that man

মীওই অসিমকনা\_DAB

meeoi asimaknaa / human-DAB → by this man

(ii) **Wh-demonstrative (DWH):** Wh-demonstratives are non-distinguishable from wh-, except for that demonstrative is followed by a noun. In Manipuri, wh-demonstratives either precede or follow a noun. However, the noun preceding the wh-demonstrative can't take nominal inflections while the one following nouns can take nominal inflections.

e.g. কৰম্বা গাড়ীনা → কৰম্বা\_DWH গাড়ীনা

karambaa gaaree-naa / which car-NOM → by which car

গাড়ী কৰম্বনা → গাড়ী কৰম্বা\_ DWH না\_NOM

gaaree karambaa-naa/ car which- NOM → by which car

**(g) Adverb (ADV):** Adverb modifies the verb with respect to place, time and manner of the action referred to by the verb. In Manipuri, adverb is formed by suffixation of -না/-na, -নো/-naa to the verbal root. Types and attributes of adverbs are- Directional, Honorific, Certainty, Inceptive, Causative, Reflexive, Commutative, Reciprocal, Evidential, Negative, Prospective, Mood, and Modality. কপ্ৰা/kapnaa, তপ্ৰা/tapnaa, তমথিনা/tamthinaa, যাংনা/yaangnaa etc. are the examples of adverb in Manipuri.

**(h) Participle (L):** A participle is a derived form of a verb which is used as a modifier of a verb. The types and attribute of participles in Manipuri are-

**Table 5.8: Different types of Participle and its attributes**

Type	Attributes
Relative Participle (LR)	Directional, Honorific, Commutative, Reciprocal, Reflexive, Evidential, Certainty, Causative, Participle Aspect, Negative, Prospective, Mood, Modality
Verbal Participle (LV)	Directional, Honorific, Commutative, Reciprocal, Reflexive, Evidential, Habitual, Certainty, Inceptive, Causative, Prospective, Definiteness
Conditional Participle (LC)	Directional, Honorific, Commutative, Reciprocal, Reflexive, Evidential, Habitual, Certainty, Inceptive, Causative, Participle Aspect, Negative, Prospective, Mood, Modality, Definiteness



(i) **Relative Participle (LR):** Relative participles in Manipuri are so formed that verbal root is followed by some verbal elements and ended by a nominalizer-

-ব/-ba or -বী/-baa or -প/-pa or -পী/-paa which is quite similar to a verbal noun. We can determine the category according to their syntactic positions. Generally, a relative participle precedes a noun phrase and hence it can't take nominal inflections. But, a verbal noun can take nominal inflections and can be the subject or object of a verb.

e.g.; ফরিবা মী → ফরিবা\_LR মী\_NC

faribaa mee / good man

(ii) **Verbal Participle (LV):** In Manipuri, the markers -দুনা, -রগা are verbal participles

e.g.; চাৰদুনা → চা\_VR র\_PROS দুনা\_LV

chaa-ra-du-na / eat-PROS-LV → by eating

চাৰগা → চা\_VR রগা\_LV

chaa-raga / eat-LV → after eating

(iii) **Conditional participle (LC):** Conditional participles are used when there are two clauses in a sentence and one clause is conditioned by the other clause so that the verb of the first clause will be in the participial form. In Manipuri, the markers -বদি/-badi, -বগা/-bagaa, -পগা/-pagaa are Conditional particles.

e.g.; লাক্ৰবদি → লাক\_VR ল\_PROS বদি\_LC

laak-la-badi/come-PROS-LC → if (someone) come

লাকপগা → লাক\_VR পগা\_LC

laak-pagaa / come-LC → just after coming

**(i) Particles (C):** Particles in Manipuri are the words which do not belong to any main category of Manipuri Part of Speeches. They are indeclinable as well as invariable in form. They have typical grammatical or pragmatic meaning. Types of particles in Manipuri are illustrated below:

**(i) Co-ordinating particle (CCD):** These particles link constituents without syntactically subordinating one to the other.

e.g.; অমসুং/amasung/and, অমদি/amadi/and

**(ii) Subordinating particle (CSB):** Subordinating particles act as conjunctions linking constructions by making one of them a complement of another.

e.g.; তৌইগুমসুং/touegumbasung/however

অদুবু/adubu/but

অদুদগী/adudagee/thereafter

**(iii) Interjection particle (CINT):** Interjections are the words which express emotion and typically fail to enter into any syntactic structures at all.

e.g.; ইস!/ish!, ইহে!/ihe!, ও!/o!, হা!/haa!, হৈমা!/heimaa!

**(j) Residual (RD):** A token that does not belong to any of the above given categories, they are to be tagged as residual. There are two different types of residuals given as below:

**(i) Foreign word residual (RDF):** The words written in any foreign script other than Manipuri is categorised as foreign word residual.

e.g.; Rocky\_RDF, Raju\_RDF, 13\_RDF etc.

**(ii) Symbol residual (RDS):** Symbols are characters which are not used as punctuation marks. They are not used as alphabets of the language also.

e.g.; Mathematical symbols, @\_RDS, #\_RDS, \$\_RDS, %\_RDS, \*\_RDS, etc.

**(k) Punctuation (PUN):** Punctuation marks are symbols that indicate the structure and organization of written language, as well as intonation and pauses to be observed when reading aloud. The markers (,), ?, !,;, %, “, etc. are punctuation markers.

**(l) Numeral (NUM):** Numerals in Manipuri are adjectives and they may precede or succede the noun.

Manipuri Numerals can be classified as follows:

**(i) Real Numeral (NUMR):** Real numerals are the numbers that can be positive or negative and have decimal places after the point.

e.g.; ১, ২, ৩, ৪, ৫, -১, -২, -৩, ২.৫ etc. / 1,2,3,4,5,-1,-2,-3, 2.5 etc.

**(ii) Calendric Numeral (NUMC):** Calendric numeral is the formatted number used to represent the date and time.

e.g.; ১২-০৩-১৯৮৮ / 12-03-1988

**(iii) Cardinal Numeral (NUMCD):** Cardinal numbers are the numbers used in counting such as the Manipuri words অমা, অনি, অহম, মরি, মঙা etc. The cardinal numeral follows the noun and there is no grammatical agreement between the noun and cardinal numeral in any construction, as in the following example:

নুপা অমা চংলি

man one go-PRG

A man is going'

Here it can be noted that there is no inflection for number to the noun নুপা 'man' and to the verb চংলি 'go + progressive' which remains same as in the above sentence.

(iv) **Ordinal Numeral (NUMO):** Ordinal numbers are the words representing the rank in a sequential order. Ordinals are formed by suffixing –সুবা to the cardinal except for the 'first' there is a separate word i.e. 'অহানবা'. The ordinal numerals in Manipuri are given below:

e.g.;	অহানবা/ahaanbaa	'first'
	আনিসুবা/anisubaa	'second'
	অহুমসুবা/ahumsubaa	'third'
	মরিসুবা/marisubaa	'fourth' etc.

In contrary of the cardinal numeral, the ordinal numeral precedes the noun as in below:

অহানবা চহী/ ahaanba chahee 'the first year'

(m) **Reduplication (RDP):** Reduplication is a morphological process in which a root or stem or part of it is repeated. Examples of Manipuri reduplicated are given below:

খুন খুন / khun khun	'villages'
যুম যুম / yum yum	'houses' etc.

### 5.3 Proposed Tagset for Manipuri

A hierarchical tagset with language specific attribute values is required to tag the lexical items of a language having large number of affixes, monosyllabic, agglutinative typological features like Manipuri. Therefore, a 3-tier tagset for Manipuri based on ILPOST framework [5] has developed. It has been

customized for Manipuri to meet the morphosyntactic requirements of the language in accordance with language specific and writing conventions followed in Manipuri.

The proposed tagset consists of 97 tags including generic attributes and language specific attribute values. The tagset can be divided into two sections viz. Tagset-I and Tagset-II. Tagset-I contains 32 tags in which 31 tags of sub categories for 13 major categories and 1 tag for unknown category. The Manipuri Tagset-I with proper examples are given below:

**Table 5.9: Manipuri Tagset-I**

Major Category	Tag	Description	Example
Noun (N)	NC	Common noun	মী ,ঈশিং ,চীঙ
	NP	Proper noun	রাজু ,চাউবা ,শিলচর
	NV	Verbal noun	চৎপা ,চেনবা ,থকপা
	NST	Spatio-Temporal	ঙশি ,হয়েঙ ,মথক ,মথা
Pronoun (P)	PPN	Personal pronoun	ঐ ,ঐহাক ,নঙ ,নহাক ,ঐখোয়
	PPS	Possessive pronoun	ঐগী ,নঙগী ,ঐখোয়গী ,ঐহাক্কা
	PDM	Demonstrative pronoun	অসি ,মসি ,অদু ,মদু
	PRF	Reflexive pronoun	ইসা ,নসা ,মসা
	PRC	Reciprocal pronoun	অমনঅমদ
	PIN	Interrogative pronoun	কনা ,করি ,কয়া ,কদায়
Verb (V)	VR	Verb Root	চৎ ,চেন ,থক
Modifier (M)	MJ	Adjective	অফবা ,অশেংবা ,ফজবা
	MQ	Quantifier	খিঙং ,খরা ,ময়াম
Specifier (SPEC)	SPEC	Specifier	মঠৈ
Demonstrative (D)	DAB	Absolute demonstrative	অদুমকনা ,অসিমকনা ,অদুদি
	DWH	Wh-demonstrative	করম্বা
Adverb (ADV)	ADV	Adverb	তপ্পা ,য়াংনা ,কম্বা
Participle (L)	LRL	Relative participle	ফরিবা
	LV	Verbal participle	দুনা ,রগা
	LC	Conditional participle	বদি ,বগা ,বদা

Major Category	Tag	Description	Example
Particle (C)	CCD	Co-ordinating particle	অমসুং ,অমদি
	CSB	Subordinating particle	তৌইগুমসুং ,অদুবু ,অদুদগী
	CINT	Interjection particle	ইস ,ইহে ,ও ,হা ,হৈমা
Residual (RD)	RDF	Foreign word residual	Raju, Exercise, 17
	RDS	Symbol residual	@, #, +, -, \$, %
Punctuation (PUN)	PUN	Punctuation	(,),!,,,ং,?,',
Numeral (NUM)	NUMR	Real Numeral	১,২,৩,৪,৫
	NUMC	Calendric Numeral	১২-০৩-১৯৮৮
	NUMCD	Cardinal Numeral	অমা ,অনি ,অহুম
	NUMO	Ordinal Numeral	অহানবা ,অনিসুবা ,অহুমসুবা
Reduplication (RDP)	RDP	Reduplication	কাক-কাক ,তে-তে ,খুন-খুন
Unknown (UNK)	UNK	Unknown	

Manipuri Tagset-II contains the morphosyntactic features or attributes of the sub categories. There are 32 attributes having 65 attribute value tags. The Manipuri Tagset -II with proper examples are given below:

**Table 5.10: Manipuri Tagset-II**

Attributes	Tag	Description	Example
Gender	MAS	Masculine gender	-অ
	FEM	Feminine gender	-ই
Number	SG	Singular number	-তোম
	DU	Dual number	-বানি
	PL	Plural number	-খোয়, -শিং
Person	1	First person	ই-
	2	Second person	ন-
	3	Third person	ম-
Case Marker	ERG	Ergative	-ন/-না
	NOM	Nominative	-ন/-না
	ACC	Accusative	-পু, -বু
	INS	Instrumental	-না
	DAT	Dative	-ঙোন্দা
	GEN	Genitive	-কি, -গি, -গী

Attributes	Tag	Description	Example
	ABL	Ablative	-দগি
	SOC	Sociative	-ক, -গ
	LOC	Locative	-দা, -তা
Nominalizer	NMZ	Nominalizer	-বা
Allative (Towards)	ALL	Allative	-শোম, -দোম
Approximate	APP	Approximate	-মুক
Similaritive	SIM	Similaritive	-কুম, -গুম, -কুম্বা, গুম্বা
Aspect	PRG	Progressive	-রি, -লি
	PRF	Perfective	-রে, -লে
Prospective	PROS	Prospective	-ল, -র
Inceptive	INC	Inceptive	-রক, -লক
Habitual	HAB	Habitual	-কন, -গন
Mood	DEC	Declarative	-ই
	SUP	Suplicative	-সি
	PROH	Prohibitive	-নু,
	IMP	Imperative	-উ, -যু
	PERM	Permissive	-সনু
	OPT	Optative	-কে, -গে
	INT	Interrogative	-ল, -র, -নো
	POT	Potential	-গ-, -ক-
	NPOT	Non-potential	-রো-, -লো-
Modality	OBL	Obligation	-তোঁ, -দৌ
	VOL	Volition	-তোয়, -দোয়
Evidential	EVI	Evidential	-লম, -রম
Certainty	CERT	Certainty	-খ
Directional	UP	Upward	-খত, -গত
	DOWN	Downward	-থ, -দ
	IN	Inward	-শিন, -জিন
	OUT	Outward	-থোক, -দোক
Causative	CAUS	Causative	-হন, -হল
Reflexive	REFL	Reflexive	-চ, -জ
Reciprocal	RECI	Reciprocal	অমনাঅমদা
Purposive	PURP	Purposive	-ন
Commutative	COMM	Commutative	-মিন
Copula	COP	Copula	-নি
Evaluative	SURP	Surprise	-নে

Attributes	Tag	Description	Example
	DUB	Dubitative	-দ, -ত
	CONF	Confirmation	-এ, -য়ে
	EXAS	Exasperation	-হে
	PERSU	Persuasion	-কো
	CMPL	Complaint	-দো
	INSIS	Insistent	-বা
Distributive	DTRB	Distributive	y/n
Definiteness	DEF	Definiteness	-তি, -দি
Emphatic	EMPH	Emphatic	-মক
Negative	NEG	Negative	-ত, -দ
Dimension	PRX	Proximal	-সি
	DST	Distal	-দু, -তু
Inclusive/ Exclusive	INL	Inclusive	-খক
	EXL	Exclusive	-তং
Honorificity	HON	Honorificity	-বি, -পি

## 5.4 Chapter Summary

In this chapter, previous work related with tagset of various languages has been discussed. It presents the different word classes of Manipuri with their grammatical constituents and their annotation procedure. Finally, the chapter presents the development of Manipuri Tagset based on ILPOST framework with a little customization to meet the morphosyntactic requirements of the language.