

Morph output standards

Amba Kulkarni

On behalf of Sanskrit Consortium

1 Morph output specifications

The morph analysis is produced as root followed by feature structure. Feature Structure is an attribute/value pair. Multiple feature structures are separated by '|'.

According to Pāṇini there are only two basic categories at the level of inflectional morphology. However, for the sake of computational purpose, we also consider avyaya as one of the categories. Later, when we would deal with the Vedic Sanskrit, we may require an additional category, upasarga.

The basic categories for morphological analysis of Sanskrit, therefore, are

- sup (noun)
- tiñ (verb)
- avyaya (indeclinable)
- upasarga (pre-position?)

Inflectional morphology

Following are the features associated with various categories.

sup: rt,vargaḥ,liṅgam,vibhaktiḥ,vacanam,level

tiñ: rt,dhātuḥ,lakāraḥ,prayogaḥ,puruṣaḥ,vacanam,padī,gaṇaḥ,sanādiḥ,level

avy: rt,level

upasarga: rt,level (Note: This category is required only for Vedic Sanskrit literature.)

The level corresponds to the level of analysis as described below:

level 1: inflectional morphology

level 2: analysis of kṛdantas

level 3: analysis of taddhitas

level 4: analysis of compounds

Derivational morphology

avytaddhita: rt,taddhita_pratyayaḥ,liṅam,level

avykṛt: rt,kṛt_pratyayaḥ,dhātuḥ,gaṇaḥ,sanādiḥ,level

kṛt: rt,kṛt_pratyayaḥ,liṅam,vibhaktiḥ,vacanam,dhāwuḥ,gaṇaḥ,kṛt_vb_rt,level

taddhita: rt,taddhita_pratyayaḥ,taddhita_rt,liṅgam,vibhaktiḥ,vacanam,level

The values of each of these features for Sanskrit is given below.

- vargaḥ
 - nā (a nāmapada)
 - sarva (a sarvanāma)
 - saṁkhyeyam (a cardinal number as an adjective)
 - saṁkhyā (a cardinal number)
 - pūraṇam (an ordinal number)
 - avy (An avyaya)

In case of derivational morphology, we indicate finer category information. Following are the subcategories of noun.

- sa-pū-pa (samāsa pūrva pada)
 - sa-u-pa-puṁ (samāsa uttarapada in pulliṅga)
 - sa-u-pa-napuṁ (samāsa uttarapada in napunsakaliṅga)
 - sa-u-pa-strī (samāsa uttarapada in strīliṅga)
 - nā_matup (derived taddhita)
 - nā_tva (derived taddhita)
 - nā_tamap (derived taddhita)
 - nā_tarap (derived taddhita)
 - nā_mayaṭ (derived taddhita)
 - nā_tal (derived taddhita)
 - nā_k-ara (derived compound)
 - nā_tṛc (derived kṛdanta)
 - nā_śatṛ (derived kṛdanta)
 - nā_ktavatu (derived kṛdanta)
- liṅgam
 - puṁ
 - strī
 - napuṁ
 - a (to indicate any possible liṅgam, e.g. in case of sarvanāma asmad)
- vacanam
 - 1 (ekavacanam)
 - 2 (dvivacanam)
 - 3 (bahuvacanam)
- puruṣaḥ
 - u (uttama)
 - ma (madhyama)
 - pra (prathama)
- vibhaktiḥ
 - 1 (prathamā)
 - 2 (dvitīyā)
 - 3 (tṛtīyā)
 - 4 (caturthī)
 - 5 (pañcamī)
 - 6 (ṣaṣṭhī)
 - 7 (saptamī)
 - 8 (sambodhana)
- lakāra

- laṭ
- liṭ
- luṭ
- lṛṭ
- loṭ
- lañ
- vidhiliñ
- āsūrlīñ
- luñ
- lṛñ
- padī
 - ātmanepadī
 - parasmaipadī
- prayogaḥ
 - kartari
 - karmaṇi
 - bhāve
- gaṇa
 - 1 (bhvādiḥ)
 - 2 (adādiḥ)
 - 3 (juhotyādiḥ)
 - 4 (divādiḥ)
 - 5 (svādiḥ)
 - 6 (tuxādiḥ)
 - 7 (ruXādiḥ)
 - 8 (tanādiḥ)
 - 9 (kryādiḥ)
 - 10 (curādiḥ)

List of dhātu_with_it will be given in an appendix.

- kṛt_pratyayah
 - tṛc
 - tumun
 - tavyat
 - yak
 - śatṛ
 - śānac
 - ghañ
 - ṇamul
 - ṇvul
 - ṇyat
 - lyuṭ
 - yat
 - ktvā
 - lyap
 - kta

- ktavatu
- anīyar
- taddhita_pratyayaḥ
 - tal
 - matup
 - tarap
 - tamap
 - tva
 - vat
 - tasil
 - karam
 - artham
 - pūrvaka
 - mayat
 - vāram
 - kṛtvasuc
 - dā
 - śas

Derivational Morphology Here are a few examples of the derivational morphology. (Output is in WX notation)

- pATayawi paT1_Nic < prayogaH:karwari >< lakAraH:lat >< puruRaH:pra>
< vacanam:1 >< paxI:parasmEpaxI >< XAwuH:paTaz > < gaNaH:BvAxiH
>< level:1 >
- Agawya Af_gam1 < vargaH:avy >< kqw_prawyayaH:lyap >< XAwuH:gamLz
> < gaNaH:BvAxiH >< level:1 >
- XanavawA Xana < vargaH:nA >< waxXiwa_prawyayaH:mawup ><
lifgam:puM > < viBakwiH:3 >< vacanam:1 >< level:3 > / Xana <
vargaH:nA >< waxXiwa_prawyayaH:mawup > < lifgam:napuM >< viBakwiH:3
>< vacanam:1 >< level:3 >
- gacCawi gam1 < lifgam:puM >< kqw_prawyayaH:Sawq >< XAwuH:gamLz
> < gaNaH:BvAxiH >< viBakwiH:7 >< vacanam:1 >< level:2 > /
gam1 < lifgam:napuM >< kqw_prawyayaH:Sawq >< XAwuH:gamLz >
< gaNaH:BvAxiH >< viBakwiH:7 >< vacanam:1 >< level:2 > / gam1
< prayogaH:karwari >< lakAraH:lat >< puruRaH:pra > < vacanam:1
>< paxI:parasmEpaxI >< XAwuH:gamLz > < gaNaH:BvAxiH ><
level:1 >