Morph User Friendly Output Format

Amba Kulkarni
On behalf of Sanskrit Consortium

June 28, 2012

1   Morph user friendly output specifications

The morph analysis is produced as a stem/root followed by a feature structure. Feature Structure is a set of values seperated by ';'s, and sometimes by ''s. 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)
- tiṅ (verb)
- avyaya (indeclinable)
- upasarga (pre-position?)

Inflectional morphology
Format for output of inflectional morphology is
sup: stem{liṅgam}{vibhaktiḥ;vacanam}
avy: stem{ "avyaya" }
tiṅ: root{prayogaḥ;lakāraḥ;puruṣaḥ;vacanam;padī;gaṇ;dhātu_with_it;sanādiḥ}
upasarga: stem"upasarga" (Note: This category is required only for Vedic Sanskrit literature.)

Derivational morphology
avytaddhita: stem{taddhita_pratyayaḥ}{liṅam}
avykṛt: root{ "kṛt_pratyayaḥ":kṛt_pratyayaḥ;dhātuḥ;gaṇaḥ}
kṛt: root{ "kṛt_pratyayaḥ":kṛt_pratyayaḥ;dhāwuḥ;gaṇaḥ;liṅam}{vibhaktiḥ;vacanam}
taddhita: stem{taddhita_pratyayaḥ}{liṅgam}{vibhaktiḥ;vacanam}

The values of each of these features is given below.

- liṅgam
  - puṁ
  - strī
  - napuṁ
  - a (to indicate any possible lifgam, 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 (tṛtiyā)
  - 4 (caturthī)
  - 5 (pañcamī)
  - 6 (ṣaṣṭhī)
  - 7 (saptamī)
  - 8 (sambodhana)

- lakāra
  - laṭ
  - liṭ
  - luṭ
  - lṛṭ
  - loṭ
  - laṅ
  - vidhiliṅ
  - āśīrliṅ
  - luṅ
  - lṛṅ

- 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)
- kṛt_pratyayaḥ
  - 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
  - mayaṭ
  - vāram
  - kṛtvasuc
  - dā
  - śas