



### 3.4 Indian Standard Lexware Format

#### Preamble

It is today an accepted fact that the lexicon is the source of power in any language processing system. The more well structured and knowledge rich the lexicon, the better is the quality of the analysis and the generation of linguistic data. The coverage of the whole gamut of the phenomena in any natural language needs constructing accurate and vast lexicons of fine quality. This is often a very large task needing quality manpower in large number.

In a country like India with multiple languages, the need for the standardization of the Lexware format was long felt so that this resource can be developed in a collaborative manner. Enabling of Indian languages on the internet and the localization of Information Technology has been taken up as a priority task by the Ministry of Communications & Information Technology (MCIT) in India. The idea is to be able to pull the resources of lexicon building activity going on at various places and in various languages. If all these efforts follow a simple and standard format, the collaborative development of this very important linguistic resource will proceed efficiently. Also the future linkages of various language lexicons will be easier.

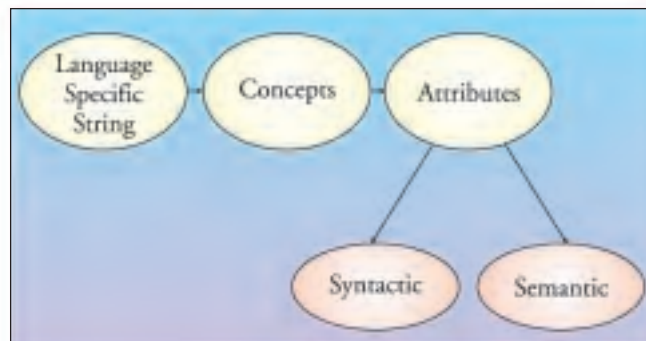
#### Basic Concepts

For any language processing system, the three most important Lexical resources are

1. Multilingual Lexicon
2. Language Word Net
3. Ontology

The multilingual lexicon defines the mapping of words from one language to another. The WordNet is a massive network which connects words with various semantic relations. The Ontology is a categorization of concepts, which is language independent, and guides the various decisions and choices in lexicon and WordNet building.

In this document we primarily discuss the standardization of the lexicon- the entities in it, the format of the entries and the linkage of the entries with the WordNet and the Ontology. The diagram in Lex. 1 illustrates the various entities that must be present in a lexicon.



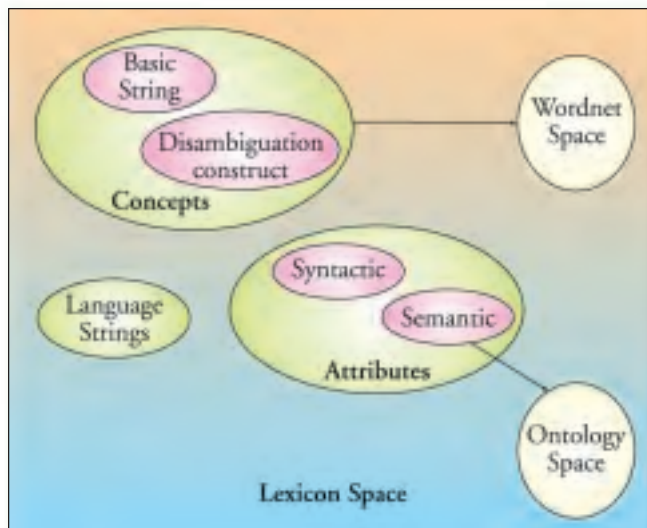
#### Lex.1 : Essential Entries in the Lexicon

Here the language specific strings map to 'concepts' which are language independent. The concepts could be represented by the words of a language after attaching disambiguation constructs to them. Thus *dog(a-kind-of animal along with other word-net attributes)* represents a language independent concept which is linked to *kuttaa* in Hindi, *kukur* in Bengali and so on. Here *dog(a-kind-of animal along with other word-net attributes)* belongs to the space of concepts while *kuttaa* or *kukur* belongs to the space of language specific strings.

The space of attributes is of great importance since the flags in the space guide the analysis and generation processes. The attributes are of two kinds- **syntactic** and **semantic**. The part of speech and the morphological behavior belong to the set of syntactic attributes, while the information like *animate/inanimate, perishable, eatable etc.* belongs to the set of semantic attributes. While the former is well defined set, the later is quite open-ended and depends on the domain and the application envisaged. However, some semantic attributes seem to be frequently used across the domains and they also can looked upon as a standard set. Lex.2 depicts the inter-relationships.

It is reemphasized in this document that our standardization effort recognizes the supreme importance of the semantic attributes. Though open ended, the lexicon building activity can be accelerated by setting up the ontology. *Thus semantic attributes connect the lexicon space with the ontology space.*

The disambiguation constructs like *a-kind-of, part-of, has-part etc.* are semantic relations with other words. *Thus the disambiguation set connects the lexicon space with the WordNet space.*



**Lex.2 : Interaction between Lexicon Space World Net Space and Ontology Space**

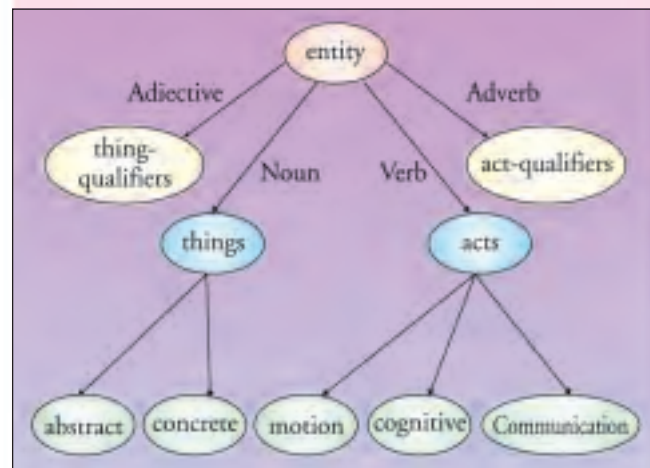
Having described these basic ideas we now proceed to the recommendation for standardization.

**Standardization**

The nomenclature used in the lexicon is the following:

- ▶ **Root Word** : Word from which the various forms are generated.
- ▶ **Morphology Paradigm Number** : An integer specifying the table in which the morphological transformation rule of the root are specified. The table specifies the strings that must be attached depending on the subject object and/or other case related information. For example, one such paradigmatic information could be *ladakA (boy)* in Hindi takes the string *oM ne* after deleting *A*, in plural number and past tense for a transitive verb.
- ▶ **Domain** : Used to indicate the domain to which a particular meaning of the word belongs. For example ~G is used for general, ~IT for information technology, ~H for health and so on.
- ▶ **English Meaning** [Example Sentence] : English Meaning of the word followed by an example sentence that implicates the meaning.
- ▶ **Syntactic Information** : This is part of speech, morphology paradigm and such other information. Also are included disambiguation rules for verbs. See below (*verb pattern scheme*).

- ▶ **Semantic Tag** : Tags or disambiguation rules for each English sense of the word. These tags are of two kinds:



**Lex.3 : A Basic Ontology**

- ▶ **Ontology Tags** : These are obtained from the directed acyclic graph (DAG) that represents the categories of the concepts. A very top-level part of this is shown in Lex.3.
- ▶ **WordNet tags** : For a particular sense of the word, the words from the wordnet which have semantic relations like *hypernym*, *meronym* etc. with the given word. For example, for *dog* we would keep *animal* (hypernym), *leg* (meronym) etc. These tags assist in uniquely identifying the meaning. We would also like to keep the sense number of the WordNet in this field.

**Verb Pattern Scheme**

Tag	Description	Example Sentence
La	Linking verb + adjective	The soup was delicious.
Ln	Linking verb + noun	Ram became a teacher.
I	Intransitive verb	Ram is sleeping.
Ipr	Intransitive verb + prepositional Phrase	People complain about the traffic.
Ip	Intransitive verb + particle	The monkeys chattered away.
In/pr	Intransitive verb + noun or prepositional phrase	The meeting lasted three hours/for three hours.
It	Intransitive verb + to-infinitive	Jane hesitated to phone the office
Tn	Transitive verb + noun	A small boy opened the door.
Tn.pr	Transitive verb + noun	The accused convinced



	+ prepositional phrase	the court of his innocence.
<b>Tn.p</b>	Transitive verb + noun + particle	The nurse shook the medicine up.
<b>Tf</b>	Transitive verb + finite 'that' Clause	Officials believe that a settlement is possible.
<b>Tw</b>	Transitive verb + wh-clause	We had not decided what we ought to do next/ what to do next.
<b>Tt</b>	Transitive verb + to-infinitive	Mary hates to drive in the rush-hour
<b>Tnt</b>	Transitive verb + noun + to-Infinitive	expect the parcel to arrive tomorrow.
<b>Tg</b>	Transitive verb +ing form of a	Peter enjoys playing football.
<b>Tsg</b>	Transitive verb + noun (+ 's) + ing form of a verb	We dread Mary/Mary's taking over the business.
<b>Tng</b>	Transitive verb + noun + ing form of a verb	She spotted a man waving in the crowd.
<b>Tni</b>	Transitive verb + noun + Infinitive	We watched the men unpack the china.
<b>Cn.a</b>	Complex-transitive verb + noun + adjective	The fridge keeps the beer cool.
<b>Cn.n</b>	Complex-transitive verb + noun +noun	The court considered Smith a trustworthy witness.
<b>Cn.n/a</b>	Complex-transitive verb + noun +as + noun or adjective	The police didn't accept the story as true (or as the fact).
<b>Cn.t</b>	Complex-transitive verb + noun -to-infinitive	The thief forced Sita to hand over the money.
<b>Cn.g</b>	Complex-transitive verb + noun + -ing form of a verb	The policeman got the traffic moving.
<b>Cn.i</b>	Complex-transitive verb + noun +infinitive	Mother won't let the children play in the road.
<b>Dn.n</b>	Double-transitive verb + noun + noun	Henri taught the children French.
<b>Dn.pr</b>	Double-transitive verb + noun + prepositional phrase	Henri taught French to the Children children.
<b>Dn.f</b>	Double-transitive verb + noun + finite 'that' clause	Colleagues told Paul that the job would not be easy.
<b>Dpr.f</b>	Double-transitive verb + prepositional phrase + finite 'that' clause	Employers announced to journalists that the dispute had been settled.

<b>Dn.w</b>	Double-transitive verb + noun + wh-clause	The porter reminded guests where they should leave their luggage/where to leave their luggage.
<b>Dpr.w</b>	Double-transitive verb + prepositional phrase + wh-clause	You should indicate to the team where they are to assemble/where to assemble.
<b>Dn.t</b>	Double-transitive verb + noun + to-infinitive	The director warned the actors not to be late.
<b>Dpr.t</b>	Double-transitive verb + prepositional phrase + to -infinitive	Fred signaled to the waiter to bring chair.

The above syntactic information are stored along with the verb entries in the lexicon to facilitate disambiguation.

**Semantic Information used for the Verbs (Used only when there are multipal meanings)**

- sth(to sb) ..... something(to somebody)
- sb(with sth) ..... somebody(with something)
- sth(from sb) ..... something(from somebody)
- sth(for sth) ..... something(for something)
- sb/sth(with sth) ... somebody/something (with something)

**Examples**

**For Noun**

- Line 1: Root Word
- Line 2: Morphology-Paradigm No *noun*
- Line 3: ~Domain1 English Meaning1 [Example Sentence]; ~Domain2 English Meaning2 [Example sentence]>; ....
- Line 4: hypernymy1[,meronymy1]; hypernymy2 [, meronymy2];....
- Line 5: ~H/VHindi Meaning1: m/f/n paradigm no, ontology tags1; Hindi Meaning2: m/f/n paradigm no, ontology tags2; ....
- Line 6: ~AS (Assamese meaning with other informations)
- Line 7: ~( Similarly meanings for other Indian Languages can be entered in separate lines one after another)

**Example for noun**

- Line 1: *head* /English root word/
- Line 2: *65 noun* /Table number with morphological and categorical information/



Line 3: ~G part of the body containing the eyes, nose, mouth and brain  
[He fell and hit his head];  
~ADM chief person of a group or organisation  
[Report to the Head immediately];  
~FIN accounts head  
[Into which head should I put the given expenditure];  
~G head of the coin  
[We tossed a coin and it came down heads]  
/ Different English senses of the root word/  
Line 4: *body\_part; post\_holder; topic; thing*  
/Different English senses of the word/  
Line 5: ~HN *sira:m 6, inanimate, concrete;*  
*aXyakRa:m 6, animate, concrete;*  
*SIrRaka:m 6, inanimate, abstract;*  
*ciwwa:m 8, inanimate, concrete*  
/Hindi meanings with gender information, paradigm type and ontology/

**Note :** It may be noted that for machine translation task, we avoid putting too many meanings in the lexical data-base unless these can be disambiguated. In this sense, it may be advisable to keep only two meanings as 'sira' and 'heda' (since the English word 'head' in all other contexts have been frequently used as it is in Hindi).

#### For Verb

Line 1: Root Word  
Line 2: Morphology-Paradigm No *verb*  
Line 3: ~Domain1 English Meaning1 [Example Sentence];  
~Domain2 English Meaning2 [Example Sentence]; .....  
Line 4: hypernymy1 [, Syntactic Info1] [, Semantic Tag1 for Subject]  
[,Semantic Tag1 for Object1] [, Semantic Tag for Object2];  
[[ ] [ ] [ ] [ ] [ ] [ ]]; .....  
Line 5: ~HN Hindi Meaning1: Paradigm-No  
[,Vibhakti Parasarg Info for Subject] [,Vibhakti Prasarg Info for Object] [, ontology tags];  
(Similar information for other meanings);...  
Line 6: ~(Similarly meanings along with the language specific information can be entered in other Indian languages in separate lines below)

#### Example for verb

Line 1: *murder* /English root word/  
Line 2: *16 verb* /Table of morphological and categorical information/  
Line 3: ~G *kill somebody unlawfully and intentionally*  
[He murdered her with a knife];.....  
/Different English senses of the root word/  
Line 4: *kill, Tn, Tn. pr; I, sb, sb- with- sth, [ ], [ ]*;  
/Disambiguation rules for disambiguating different senses of the verb/  
Line 5: ~HN *hawtyaA kara:257 ne, ko, VoA*;/Hindi meaning with subject and object  
Vibhakti parasarga and ontology tags/

#### For Adjective

Line 1: Root Word  
Line 2: Paradigm-No *adjective*  
Line 3: ~Domain1 English Meaning 1 [Example Sentence];  
~Domain2 English Meaning 2 [Example Sentence]; .....  
Line 4: hypernymy1[, antonymy1]; hypernymy2 [,antonymy2];.....  
Line 5: ~HN Hindi Meaning1, ontology tags1; Hindi Meaning2, ontology tags2; ...  
Line 6: ~ (Similarly meanings in other indian Languages can be entered in lines below)

#### Example for adjective

Line 1: *mysterious* / English root word /  
Line 2: *12 adjective* /Table of morphological and categorical information/  
Line 3: ~G *difficult to understand or explain* [She gave me a mysterious look]/Different English senses of the root word/  
Line 4: *state, clear* /Wordnet Tag/  
Line 5: ~HN *rahasyamaya*, Hindi Meaning/

#### For Adverb

Line 1: Root Word  
Line 2: Morphology-Paradigm No *adverb*  
Line 3: ~Domain English Meaning1 [Example Sentence];  
~Domain English Meaning2 [Example Sentence]; .....  
Line 4: hypernymy1[, synonymy1]; hypernymy2[, synonymy2];.....  
Line 5: ~HN Hindi Meaning1[, ontology tags1];



Hindi Meaning2[, ontology tags2];.....

Line 6: ~ (Similarly meanings in other indian Languages can be entered in lines below)

### Example for adverb

Line 1: *mysteriously* /English root word/

Line 2: *13 adverb* /Table of morphological and categorical information /

Line 3: *~G mysteriously [The main witness had mysteriously disappeared];*  
/Different English senses of the root word/

Line 4: *manner* /WordNet Hypernymy/

Line 5: *~HN rahasyamaya DaMga se,* /Hindi Meaning/

### Semantic Attributes For Lexware Standard

#### Noun

- Proper Noun (PROP, eg, rAma)
- Common Noun (If a noun is neither Proper nor Collective then it is taken for granted that it is Common Noun and hence, no symbol is kept for it.)
- Collective Noun (COLCT, eg, BidZa)
- Animate (ANIMT)
  - Flora (FLORA)
    - => Shrubs (FLORA-SHRB , eg, tulasI)\*
    - => Aquatic plants (FLORA-AQTC, eg kamala)
    - => Climbers (FLORA-CLMB, eg , aMgura kI bela)
    - => Trees (FLORA-TREE, eg, Ama)
  - Fauna (FAUNA)
    - => Mammals
      - Person (ANIMT-FAUNA-MML-PRSN, eg ladZakI)
      - Ape (ANIMT-FAUNA-MML-APE, eg, laMgUra )
      - Lesser Mammals (ANIMT-FAUNA-MML-LSMML, eg dAzlPZina)
    - => Reptiles (ANIMT-FAUNA-RPTL, eg, sAMzpa)
    - => Amphibians (ANIMT-FAUNA-AMPHB, eg, kaCuA)
    - => Aquatic Animals (ANIMT-FAUNA-AQAN, kekadZA)

- => Birds (ANIMT-FAUNA-BIRD, eg, wowA)
- => Fish (ANIMT-FAUNA-FISH, eg, SARka)
- => Insects (ANIMT-FAUNA-INSCT, wiwalI)
- => Micro organism (ANIMT-FAUNA-MCORG, eg, bEktIriyA )
- => Imaginary Animals (ANIMT-EAUNA-IMGYAN, eg, drEgana)

#### ● Inanimate (INANI)

##### ■ Object

- => Artifact (INANI-ARTFCT, eg, cammaca )
- => Natural Object (INANI-NAT-OBJCT, pahAdZa)
- => Edible (INANI-EDBL-OBJCT, eg, miTAI)
- => Anatomical (INANI-ANTM-OBJCT, eg uMgalI, bAla)
- => Chemical (INANI-CHML-OBJCT, eg amla)
- => Physical (INANI-PHSCL-OBJCT, eg kalama, maMca)
- => Imaginary (INANI-IMGN-OBJCT, eg amQwa)

#### ● Place

- => Imaginary Place (INANI-IMGY-PLC, eg, svarga)
- => Physical Place (INANI-PHSCL-PLC, eg, pATaSAIA)

#### ● Event

- => Natural Event (INANI-NAT-EVENT, eg, BUkaMpa)
- => Historical Event (INANI-HIST-EVENT, eg, praWama viSvayuxXa)
- => Planned Event (INANI-PLND-EVENT, eg, bama-visPota)
- => Social Event (INANI-SCL-EVENT, eg, janma)
- => Fateful Event (INANI-FTFL-EVENT, eg, lAztaRI nikalanA)
- => Fatal Event (INANI-FTL-EVENT, eg, xurGatanA)

#### ● Abstract (ABS)

- => Quality (INANI-ABS-QUAL, eg, acCAI)
- => Perception (INANI-ABS-PRCP, eg, sUcanA)



- => Cognition (INANI-ABS-COGN, eg kalpanA')
- => Colour (INANI-ABS-COLR, eg, lAla)
- => Title (INANI-ABS-TITL, eg, proPZesara)
- => Measurement (INANI-ABS-MSRMNT, eg, lambAI)
- => Time (TIME)
  1. Period (INANI-ABS-TIME-PRD, eg, GaMtA,)
  2. Season (INANI-ABS-TIME-SSN, eg, garmI)
  3. Historical ages (INANI-ABS-TIME-HIST, eg, pARANa yuga)
- => Action
  1. Social (INANI-ABS-ACT-SCL, eg, vivAha)
  2. Anti-social (INANI-ABS-ACT-ANTISCL, eg, corI)
  3. Occupation (INANI-ABS-ACT-OCP, eg, axyApana)
  4. Communication (INANI-ABS-ACT-COMM, eg, parAmarSa)
  5. Physical Action (INANI-ABS-ACT-PHSCLACT, eg, dubakI)
- => Object (INANI-ABS-OBJCT, eg SabXa)
- => Logos
  1. Religion (RLGN)
  2. Philosophy (PHIL) (Metaphysics, Epistemology, Logic, Ethics)
  3. Social Sciences (SCLSC) : political science, economics, commerce, law, public administration, social services, education, anthropology, psychology, sociology, folklore.
  4. Language (LNG) (Grammar)
  5. The Arts (ARTS) : Fine and Performing Arts (Music, Literature, Painting, Sculpture, Film, Drama), Aesthetics and Rhetoric, Useful arts & Crafts
  6. Geography (GEOG)
  7. History (HIST)
  8. Natural Sciences (NATSC) (Physics, Chemistry, Bio-sciences, Astronomy, Geology)
  9. Mathematics (MATHS)
  10. Applied Sciences (APPSC): Engineering,

- Agriculture, Medicine & Health, Manufacturing, Building & Construction, Ecology
- 11. Sports (SPRT) (Indoor games, Outdoor games)
- 12. Transport (TRANSPT)
- 13. Home Science (HSC) (Food & Nutrition)
- 14. Mass media (MSMDA) (Journalism, Advertising)
- 15. Fashion Designing (FSHD) (Dresses, Textile)
- 16. Miscellaneous (MSL)

- State (STE)
  - => Physical State (INANI-STE-PHSCL)
    1. Solid (INANI-STE-PHSCL-SLD, eg, pawWara)
    2. Liquid (INANI-STE-PHSCL-LQD, eg, xUXa)
    3. Gas (INANI-STE-PHSCL-GAS, eg, AzksIjana)
  - => Disease (INANI-STE-DIS, eg, buKara)
  - => Biological State (INANI-STE-BIO, eg, bacapana)
  - => Mental State (INANI-STE-MNTL, eg, avasAxa)
  - => Social State (INANI-STE-SCL eg, haMgAmA)
- Process (PRCS)
  - => Physical Process (INANI-PHYS-PRCS eg KANa banAne kI viXi)
  - => Mental Process (INANI-MNTL-PRCS eg yojanA)

#### Miscellaneous Noun Attributes :

- Abbreviation (ABRV, eg, UNL)
- Acronym (ACRNM, eg, UNESCO)
- Heading (HEAD, eg, THE INTERNATIONAL TELECOMMUNICATION UNION)

#### Verb

- State (STE)
  - => Physical State (INANI-STE-PHSCL)
- Verb of Action (VOA)
  - Change (VOA-CHNG, eg, baxalanA)
  - Cognition (VOA-COGN, eg, vicAra karanA, nirNaya lenA)



- Commencement (VOA-CMNCT, eg, barasane laganA)
- Communication (VOA-COMM, eg, liKanA, bolanA)
- Competition (VOA-COMPT, eg, ladZanA)
- Completion (VOA-CMPLT, eg, KA cukanA)
- Consumption (VOA-CNSMP, eg, KanA, pInA)
- Contact (VOA-CNTCT, eg, CUnA)
- Creation (VOA-CRTE, eg, banAnA)
- Destruction (VOA-DSTN, eg, cUra-cUra karanA)
- Emotion (VOA-EMOT, eg, haMzsanA)
- Event (VOA-EVNT, eg, barasanA, himapAwa honA)
- Grooming (VOA-GROOM, eg, saMzvaranA)
- Maintenance (VOA-MNTC, eg, yaWA-sWiwi banAe raKanA)
- Motion (VOA-MOTN, eg, calanA)
- Perception (VOA-PRCP, eg, xeKanA, sunanA)
- Performance (VOA-PRFM, eg, nQwya karanA)
- Possession (VOA-POSS, eg, kabjZA karanA)
- Social (VOA-SCL, eg, cunAva ladZanA, SAxI karanA)
- Verb of State (VOS)
  - Physical State (VOS-PHY-ST, eg, KadZA rahanA)
  - Mental State (VOS-MNT-ST, eg, ciDZanA)
- Temporal Verbs (TMP, eg calanA)
- Verbs of Continuity (CONT, eg bahanA)
- Verbs of Volition (VLTN, eg KAna)
- Verbs of Non-volition (NVLTN, eg vivaSa honA)

#### Special Verb Attributes :

- Idiom (V-IDM, eg, cUhe billi kA Kela karanA)

#### Adjective

- Descriptive
  - Weight (ADJ-DES-WT, eg, BAri puswaka)
  - Shape (ADJ-DES-SHP, eg, laMba rAswa)
  - Colour (ADJ-DES-CLR, eg, lAla kapadZA)
  - Strength (ADJ-DES-STRNGTH, eg, kamajZora kadZI)
  - Qualitative (ADJ-DES-QUAL, eg, acCA ladZakA)

- Appearance (ADJ-DES-APPR, eg, suMxara ceharA)
- Speed (ADJ-DES-SPD, eg, XimI cAla)
- Depth (ADJ-DES-DPTH, eg, gaharA jfAna)
- Existence (ADJ-DES-EXST, eg, upasWiwa loga)
- Numeral (ADJ-DES-NUM, eg, pAMzca uMzgaliyAMz)
- Temperature (ADJ-DES-TEMP, eg, garama xUXa)
- Quantitative (ADJ-DES-QUAN, eg, WodZA pAnI)
- Respective (ADJ-DES-RESP, eg, SrImawI iMxirA gAMXI)
- Emotion (ADJ-DES-EMOT, eg, kroXIwa vyakwi)
- Demonstrative (ADJ-DMON, eg, yaha ladZakA)
- Interrogative (ADJ-INTRO, eg, kisakA makAna hE?)
- Relational (ADJ-REL, eg, mOserA BAI)

#### Special Adjective Attribute :

- Nouns used as Adjective (N-ADJ, eg, kaMpyUtara Kela, majaxUra saMGa)

#### Adverb

- Time (ADV-TIME, eg, ke pahale)
- Frequency (ADV-FREQ, eg, bAra-bAra, xo bAra)
- Place (ADV-PLC, eg, hara jagaha)
- Manner (ADV-MAN, eg, wejZI se)
- Quantity (ADV-QUAN, eg, bahuwa)
- Reason (ADV-RSN, eg, isalie)
- Interrogative (ADV-INTRO, eg, kaba)
- Affirmative (ADV-AFRM, eg, niSciwa hI)
- Negative (ADV-NGTV, eg, nahIM, SAYaxa)

#### Conclusions

The above discussion forms the basis of the standardization work and is called the **Lexware Standard Document: (Foundations)**. This has been evolved from the experiences of the various lexical resources development activity going on at IIT Kanpur, IIT Bombay and such other places. This document will be followed by:



- I. Lexware Standard Document: (Morphology Paradigm Tables).
- II. Lexware Standard Document: (Domain Categorization).
- III. Lexware Standard Document: (Lexware Developers' Manual).

The main aim for all this effort is to ensure that a massive multilingual lexical resource is built for the languages of India. This will ultimately lead to (semi) automatic MT systems with capability for NL understanding. As mentioned before, this is a manpower intensive activity requiring well trained personnel in large number. The task has to be executed in a distributed manner. It is envisaged that each resource center will establish linkages to the language of the center as shown in the examples above.

Transliteration scheme followed in the above description is given below :

अ a, आ A, इ i, ई I, उ u, ऊ U, ऋ q, ए e, ऐ E, ओ o, औ O, ँ M, ः H, ण z, क k, ख K, ग g, घ G, ङ f, च c, छ C, ज j, झ J, ञ F, ट t, ठ T, ड d, ढ D, ण N, त w, थ W, द x, ध X, न n, प p, फ P, ब b, भ B, म m, य Y, र r, ल l, व b, श S, ष R, स s, ह h, क्ष kR.

**Example :** hiMxI akRaramAIA = हिन्दी अक्षरमाला

**Note :**

1. A character with nukta is represented by corresponding Devanagari symbol followed by 'Z' e.g. ज -> jZ. 'Z' is "nukta operator".
2. For other nondevanagari symbols we use 'V' & 'Y' as "previous" & "following" operators respectively e.g. ऐ -> eV, ऌ -> lY.

It is expected that this document will be carefully read, discussed and reflected upon. Also some languages may have their special requirements. Any feedback or comment pertaining to this may be sent to one of the following:

(Prof. R.M.K. Sinha, [rmk@cse.iitk.ac.in](mailto:rmk@cse.iitk.ac.in), Tel: 0512-597174  
Prof. Pushpak Bhattacharya, [pb@cse.iitb.ac.in](mailto:pb@cse.iitb.ac.in), Tel: 022-5767718  
Prof. Uma Maheswar Rao, [guraosh@uohyd.ernet.in](mailto:guraosh@uohyd.ernet.in))

### Proposed Standard for Indian Script to Roman Transliteration Table (INSROT)

Indian Scripts are phonetic and have similarity in alphabetic correspondence. Transliteration between Indian languages is simple, unambiguous and phonetically similar. However, there is need for a Scheme for transliteration from Indian Scripts to Roman Script. There have been schemes but they lack in readability of the transliterated text.

INSROT (Indian Script Roman Transliteration Table) is proposed as a standard transliteration scheme considering the need for readability, and ease in de-coding the romanized text unambiguously. This is orthographic representation. Transliteration table, rules of syllable formation & disambiguation/decoding and examples are given below:

#### Vowels

Syllabic Form	Intra Syllabic Form	INSROT
अ	-	a
आ	ा	A
इ	ि	i
ई	ी	I
उ	ु	u
ऊ	ू	U
ए	े	e
ऐ	ै	E
ओ	ो	o
औ	ौ	O
अं	ँ	ah
अँ	ं	Ah
ऋ	ृ	Ri
ऌ	ृ	RI
ऍ		IRi
ऎ		IRI
ं		◌M
ँ		◌Mh
ः		:

=> The character ऋ (Ri) / ऌ (IRi) does not represent a single vocalic sound in Hindi, but is vocalic in terms of the Script, having separate syllabic and intra syllabic forms.





### Consonants

	Voiceless Unaspir. अघोष अल्पप्राण	Voiceless aspirated अघोष महाप्राण	Voiced unaspir. घोष अल्पप्राण	Voiced aspirated घोष महाप्राण	Nasals नासिक्य नासिक्य
Velars (कंठ्य)	क (ka)	ख (kha)	ग (ga)	घ (gha)	ङ (Nha)
Palatals (तालव्य)	च (cha)	छ (chha)	ज (ja)	झ (jha)	ञ (nha)
Retroflexes (मूर्धन्य)	ट (Ta)	ठ (Tha)	ड (Da)	ढ (Dha)	ण (Na)
Dentals (दन्त्य)	त (ta)	थ (tha)	द (da)	ध (dha)	न (na)
Labials (औष्ठ्य)	प (pa)	फ (pha)	ब (ba)	भ (bha)	म (ma)
Semi vowels (अर्ध स्वर)		य (ya)			व (va)
Liquid (तरल)			र (ra)	ल (la)	
Sibilants (संघर्षी)	श (sha)		ष (Sa)	स (sa)	
Glottal (काकल्य ध्वनि)			ह (Ha)		
		क्ष (kSa)	त्र (tra)	ज्ञ (jnha)	

### Nukta Consonants

क् (k.a)	ख् (kh.a)	ग् (g.a)	ज् (j.a)
ड् (D.a)	ढ् (Dh.a)	फ् (ph.a)	र् (r.a)
व् (v.a)			

### Explicit Halant

् =X (not-join), पक्व ~ pakva, पक्व ~ pakXva,  
पंक~ paMka, पङ्क ~ paNhkA, पङ्क ~ paNhxka

### Intra Syllabic Vowels Forms combined with consonant character

क (ka)	का (kA)	कि (ki)	की (kI)
कु (ku)	कू (kU)	कृ (kRi)	के (ke)
कै (kE)	को (ko)	कौ (kO)	कॉ (kAh)
कं (kaM)	कँ (kaMh)		

### Rule for decoding from Romanized text

'a' is integral part of consonant. Omission of 'a' results into 'pure' consonants. e.g. ka (क), k (क् / क).

Vowel combined with 'pure' consonant transforms into mAttrA. e.g. khi (खि), ghU (घू), kAh (कॉ).

Full stop "." should be followed by a space.

Capital letters denotes 'dIrgh/prolonged vowel/ mAttrA. e.g. [ a (अ), A (आ)], [ i (इ), I (ई)], [ ku (कु), kU (कू)], [ ke (के), kE (कै)], ...

h is added to consonants/vowels to denote aspirated or closer sounds, eg. kha (ख), chha (छ), jha (झ), ...

Nha (ङ), nha (ञ), ah (अँ), Ah (आँ), aMh (अँ), AMh (आँ), kaMh (कँ)

M denotes anusvAr/bindu and combines with previous vowel as in

aM (अं) , kaM (कं) , kiM (किं), sviMga (स्विंग)

अँ = aMh ; ँ = ँMh

[Note: Graphemic code joining as अँ + ँ = अँ (ahM) is not permitted.]

eg. kh (ख), Nh (ङ), Ah (आँ), ँ (ँMh), हँ (HaMh)

### Examples of word/phrase level transliteration

कट (kaTa), खाट (khATa), घृत (ghRita), चिप (chipa),  
झील (jhIla), बाण (bANa), तुक (tuka), थूक (thUka),  
फोड़ (phoD.a), मैल (mEla), षट (SaTa), साथ (sAtha),  
आशा (Asha), ईख (Ikha), ऋषि (RiSi), और (Ora),  
बूढ़ा (bUDh.A), मक्खन (makkhana), ड्राइवर  
(DrAivara), अण्डा (aNDA), गद्दी (gaddI), सत्ताईस  
(sattAIsa), पृथ्वी (pRithvI), प्रथम (prathama), कुँवर  
(kuMhvara), ख़ाब (kh.bAb), श्वान (shvAna), प्राण  
(prANa), पुण्य (puNya), गङ्गा (gaNhgA), ज्ञानज्योति  
(jnhAnajyoti), नूपुर (nUpura), ध्वनि (dhvani), सुनि  
(suni), कॉलेज (kAhleja), हंस (HaMsa), हँस,  
(HaMhsa), आँखें (AMhkheM), विपत्ति (vipatti),  
परिस्थिति (paristhiti), पॉन (pahna), बॉल (bAhla), बाँके  
बिहारी (bAMhke biHarI), गड्गा (gaNhXgA), क्वाथ  
(kXvAtha), क्वा (kvA), पद्य (padya), पद्य (padXya),  
पूर्वी (pUrvI), ट्रेन (Trena), प्राण (prANa).

[Suggestions on the above draft

INSROT are requested.

Contact : [tdilinfo@mit.gov.in](mailto:tdilinfo@mit.gov.in)]