**UNIVERSAL SEMANTIC REPRESENTATION GUIDELINE VERSION 4.2**

[**Motivation of USR**](#_qmidieiwtmjc) **1**

[Convention of symbols used in USR](#_e3deks85zpwz) 2

[**Format of USR**](#_v74lbrpueyxr) **3**

[**Sentence Segmentation**](#_75buigh40wu5) **3**

[**Row 1: Original Sentence**](#_tglqjkgi67w2) **7**

[Sub-sentence Identification](#_iznnf3eju7rx) 7

[**Row 2: Concepts**](#_jmjjf1ll3c5b) **9**

[What to present in the concept row?](#_5zebl82zdzg4) 9

[How are concepts represented?](#_139h4xzdv10c) 11

[Why Concept ID?](#_uujoryk6ikba) 11

[Note](#_ksp10kfcdqlw) 13

[Which items are not presented in the concept row](#_lhuagt5uqs9k) 13

[Entity](#_adojazh23zed) 14

[Pronominal Reference to an Entity](#_9sf1jgui85k3) 14

[Personal Pronoun](#_e52z6o6a0bk9) 14

[How to distinguish among tuma, tū and āpa (the three forms of 2nd person pronoun)](#_i2tq8nmshqql) 14

[wyax](#_ywxwys4adunr) 15

[How to distinguish between proximal and distal form of wyax](#_1b1jo58z2cfp) 15

[Reciprocal Pronouns](#_eeexwtmtm5f4) 15

[Reflexive Pronoun](#_l0ivhvma1tl7) 15

[Interrogative Pronoun](#_uulut4m3c8sy) 15

[Relative Pronoun](#_s16a9gmvnna) 15

[Example for all Pronouns](#_h0roo9630109) 15

[Events](#_klo15gty9m5s) 16

[Concepts for verb forms consist of root and TAM info separated by ‘-’. `](#_evi3wn7by85w) 16

[Tense, Aspect, Modality (TAM)](#_6eibzdw6xlcj) 17

[Modifiers of Entities](#_698803ilwuxe) 17

[Modifiers of Events](#_op73t0xri2k4) 18

[**Row 3: Index for the Concepts**](#_ou91u5df1b4t) **19**

[**Row 4: Semantic Categories of Nouns**](#_2viyvzwv5a6j) **19**

[Example of Semantic Category Row](#_ssp2uabl1dex) 20

[**Row 5: Morpho-Semantic Information**](#_gplu1kzid43k) **21**

[General Information](#_fdssu4ny7zyd) 21

[**Row 6: Dependency Relation**](#_88pfb85s2vkb) **22**

[**In this row we mark the relation between head and dependent in following way-**](#_hvlj3mv5388m) **22**

[kāraka relations](#_ht6sqa4ftbxx) 23

[exemplifyOther kāraka relations](#_dzjffyb6em81) 24

[Relations associated with kāraka](#_ejdf32tdh4nt) 24

[kāraketara relation](#_5gscvs7iy587) 25

[samānādhikaraṇa](#_4uwes27ioorh) 25

[bhāvalakṣaṇa:](#_aafktufr1zyn) 26

[kālavācī](#_rly1tzearuqq) 26

[Spatio-temporal Information](#_9p26szwg6gqa) 26

[Intra-sentential sangati](#_63dr01trvx59) 27

[Genitive or Possessive relation between two entities](#_38mffdmwf72k) 27

[Sādrisya, vibhājana and nirdhāraṇa](#_4nualgwcg56h) 27

[Other kārakatera relation](#_vzq7d62cskur) 28

[Modifier and modified relations](#_i59wnrsq9dun) 29

[Different Measurement relations](#_n9h2cvq6zrj0) 29

[**Row 7: Discourse Elements**](#_7krmiwx41x5x) **30**

[kAryakAraNa Discourse Connective Relation](#_86q3x9hey60x) 30

[Pronominal coreference](#_5bcv7oxrdwgu) 32

[**Row 8: Speaker’s view**](#_5d12jaw9hs6j) **33**

[Proximal and Distal Information for wyax](#_qc3blhhzii6f) 34

[**Row 9: Scope**](#_o4odlczd3uio) **35**

[**Row 10: Sentence Types**](#_c1demvv8dlno) **35**

[**Row 11: Construction**](#_njxsdgqhamsu) **36**

[**Phenomena**](#_sdjkrwnqbmjo) **38**

[Spatio-directional Terms](#_xlvzfnkyi075) 38

[Measuring Terms](#_4r6u8u8vefw5) 40

[Demonstrative](#_mbefu02bzz0o) 40

[Interrogative](#_qfztoz1cqnbi) 41

[Complex Predicate](#_1jbwqj8pyss4) 42

[Light verbs](#_dyqe4psejj8k) 42

[Complement Clause](#_8t3vsrcjgm0) 43

[Relative Clause](#_1nkrw3klefe3) 44

[**Appendix-1**](#_4jb4zeh1q42w) **51**

[Devanagari-WX-Indic script mapping](#_fotct3z7lfkp) 51

[**Appendix-2**](#_h4olo5nsx827) **52**

[Concept Dictionary Entry Format](#_11l8oxxlji8v) 52

[**Appendix - 3**](#_ljybm050mn7q) **52**

[**Appendix-4**](#_r2chpsyqtke1) **52**

[List of Hindi interrogative pronoun](#_jkxce3ayzz7n) 52

[**Appendix - 5**](#_4urrgwfvxh7o) **53**

[Hindi TAM Dictionary](#_wya5z716yvfp) 53

[**Appendix- 6**](#_w5lrt3rmh8th) **55**

[List of Spatio-directional terms in Hindi](#_akewxxjwbhwh) 55

[**Appendix-7 List of Discourse Connectives**](#_2nuputfaxut6) **56**

[**Appendix-8**](#_uwnyvekrwq0d) **60**

[List of Hindi Discourse Particle and evaluative dimensions](#_w7mktrkhdnq9) 60

[**Appendix 9**](#_g07b95ra7vn8) **68**

[List of shade or light verbs](#_ama1jvn5gzya) 68

[**Appendix-10**](#_z3kc9fwec0u) **69**

[Construction Frame in USR](#_6lm3atttnmin) 69

[Sample USR](#_imn4uv9k5nk2) 74

[USR: A brief outline](#_kailq8lzxofk)

Universal Semantic Representation (USR) is a meaning representation that models Indian Grammatical Tradition (IGT). Meaning (or thought) is there in the mind of the speaker (author) and while speaking (writing), (s)he makes use of language (or linguistic expressions) to express his/her thought. Thus a discourse (text) represents the speaker's thought.

This guideline is created to help annotators to make USRs of the written discourse. The objective is to generate multiple natural languages from these USRs using Natural Language Generators.

# Motivation of USR

Written text expresses the speaker’s intention of how (s)he wants to depict a situation. A situation can be seen as an event with various participants involved in that event and also various associated events either causing or resulting or associating with the main event. For example, let us take a situation where the main event is *speaking*. Two participants involved are *Ram* and *Sita* in the role of speakers. The location of the event is *bus stop*. The target person speakingis the *brother* of *Ram* and *Sita*. When the speaker wants to talk about this situation (s)he has to choose a tense and aspect. For example, the chosen time is past perfective in this case. This very basic situation (which we can call 'propositional information') can be expressed in Hindi as

1. rāma aura sitā ne basa aḍḍe para bhāī ke sātha bāta kī.

Now, the speaker wants to add some more situational information on this basic propositional information. They are the following: the *brother* is younger to *Ram* and *Sita*. The speaker wants to negate the whole situation. In Hindi, the chosen linguistic elements are negation marker *nahīṃ* ‘not’. Moreover, the speaker wants to add the information of certainty to the negation of the above situation. However (s)he wants to leave open the possibility of the agents’ speaking to somebody else in the bus stop. Such intention of the speaker can be expressed through the discourse particle *to* in Hindi*.* Thus the sexemplifyentence generated in different languages including Hindi is as below-

| **Language** | **Expected Outcome of the Sentence** |
| --- | --- |
| Hindi | rāma aura sitā ne basa aḍḍe para apane choṭe bhāī ke sātha to nahīṃ bāta kī. |
| Bangla | rāma āra sitā bāsa sṭaiṃḍ-e nijera choṭa bhāīera sāthe to kathā bal-e ni. |
| Nepali | rāma ra sitā-le basa-bisaunī-mā āphno sāno bhāī-samga ta kurā gare-nan |
| Telugu | rāma sitā basa sṭaiṃḍ-lo vāīyīya cinna tammu-du-to ayite mātlāda ledu |
| Punjabi | rāma te sitā apane vīra nāla te basa sṭaiṇḍa to gala ni karyā |
| Marathi | rāma āñi sītene basasthāNakāvara apalyā choṭyā bhāvāSi tar nāhī bolale. |
| Tamil | rāma un Sita vum nichayama avunga thambi kitta pesavaeilla |
| English | Ram and Sita did not certainly talk to their younger brother at the bus stop. |

Table 1. Example of expected generated sentences in different languages from a given USR

There can be one more interesting interplay of negation and certainty information in this case. The speaker here wants to say that (s)heis certain that Ram and Sita did not talk to their younger brother in the bus stop. Thus certainty takes a wider scope on negation of the actual event of Ram and Sita’s speaking with their younger brother at the bus stop. Instead, if the speaker wanted to express that he is not certain if Ram and Sita spoke to their younger brother in the bus stop, then the semantics of negation *nahīṃ* would take the wider scope over *to* ‘expressing certainty’.

In both cases, the sentence generated would have been the same. However, in USR, we have the opportunity to specify the scopal information. The speaker can annotate the appropriate scopal order of negation and discourse particles to express what (s)he actually means.

A text contains a series of sentences. Sometimes, the relation among the sentences are explicitly marked through discourse markers. These discourse markers maintain the flow of the story. For example, the speaker in this case might want to justify why (s)he assumes that *Ram* and *Sita* did not speak with their younger brother that day. In order to express that thought, the sentence generated can be:

1. Hindi: kyoṃki usa dina unakā bhāī śahara meṃ thā hī nahīṃ

Bangla: kāran sedina oder bhāi sahar-e chi-lo-i nā

*kyoṃki* ‘because’ is a discourse connective marker that logically connects (1) and (2) by justifying (1) through (2). *usa*- and *una*- (pl of *usa*-) are anaphoric pronouns. *usa dina* refers to the same day when the event took place. *una* in *unake* *bhāī* refers to *Ram* and *Sita*. These anaphoric expressions are the mechanism for maintaining the cohesiveness in the story. The discourse particle *hī* again like *to* in (1) add extra-propositional meaning which actually conveys the speaker's view or perspective.

USR attempts to capture all this information in a human-friendly yet machine tractable representation.

### Convention of symbols used in USR

| Concept and Rows | Symbol | Example |
| --- | --- | --- |
| Original sentence | # | #और आधारभूत संकल्पनाओं के साथ-साथ तकनीकी शब्दों की व्याख्या करता है, जो भौगोलिक ज्ञान के घटक हैं । |
| Sentence Type | % | %affirmative %imperative etc |
| Construction | \* | \*conj \* span etc |
| pronouns | $ | $speaker, $addressee, $wyax, $yax, $kim, |
| Foreign word | ^ | ^word\_1 |
| Abbreviation | @ | @eic.sl.yU., @nAsA |

# Format of USR

The meaning is represented in 11 rows in csv (comma (,) separated value) format. This document guides the annotators to annotate each row. The 11 rows are:

| Row 1 | Original Sentence |
| --- | --- |
| Row 2 | Concept |
| Row 3 | Index |
| Row 4 | Semantic Category of Nouns |
| Row 5 | Morpho-Semantic Information |
| Row 6 | Dependency Relation |
| Row 7 | Discourse Element |
| Row 8 | Speaker’s View |
| Row 9 | Scope |
| Row 10 | Sentence Type |
| Row 11 | Construction |

Table 2. Rows of USR

# Sentence Segmentation

Since USR annotation of complex sentences is difficult and automated USR generation for complex sentences is a challenge as observed through several experiments, we have decided to first segment complex sentences into discourse units without losing information. Some complex sentences are not segmented as segmenting them will make the discourse less coherent.

Following are the strategies of sentence segmentation

* In general, segmented segments will be a discourse unit which contains a finite verb.
* A discourse unit is a simple sentence or a clause which is not necessarily the smallest unit. It participates in making the larger discourse.

Such as- rāma aura sitā ne basa aḍḍe para bhāī ke sātha bāta kī. ‘Ram and Sita spoke to their brother in the bus-stand.’

* Relative Clauses with the relative pronoun referring to a noun in the sentence are not segmented. Such as -

bhārata kā sabase dakṣiṇī biṃdu jo iṃdirā biṃdu kahā jātā thā, san‌ 2004 meṃ jalamagna ho gayā.

‘The southernmost point of India, which was known as Indira point, was submerged in water in the year 2004.’

This sentence is not split.

**When to split Relative Clauses**:

1. If a sentence contains more than one relative clause, relative clauses are segmented and their inter-clausal relations are shown in discourse element row. Such as-

| Sent\_ID\_1 | pṛthvī ke dharātala ke ūṁce uṭhe hue bhāga jinakā śikhara hajāra mīṭara se adhika ūṁcā ho aura ḍhāla tīvra ho, tathā jinake banane jinakā lākhoṃ varṣa lage, parvata kahalāte haiṃ| |
| --- | --- |

The above sentence contains more than one relative clauses and they will be segmented as following

-

| Sent\_ID\_1a | pṛthvī ke dharātala ke ūṁce uṭhe hue bhāga parvata kahalāte haiṃ |
| --- | --- |
| Sent\_ID\_1b | jinakā śikhara hajāra mīṭara se adhika ūṁcā ho |
| Sent\_ID\_1c | aura jinakā ḍhāla tīvra ho |
| Sent\_ID\_1d | tathā jinakā banane me lākhoṃ varṣa lage |

See [Relative Clause](#_1nkrw3klefe3) for annotation rules

1. If a relative pronoun functions as a discourse connective, those relative clauses will be splitted. Such as,

nadī ke nicale bhāgoṃ meṃ ḍhāla kama hone ke kāraṇa nadī kī gati kama ho jātī hai, jisake pariṇāmasvarūpa nadīya dvīpoṃ kā nirmāṇa hotā hai.

Here, the whole expression jisake pariṇāmasvarūpa acts as a discourse connective. Hence, the clause it is attached with, is splitted from the previous clause it is connecting with and the two sentences will be:

nadī ke nicale bhāgoṃ meṃ ḍhāla kama hone ke kāraṇa nadī kī gati kama ho jātī hai.

isake pariṇāmasvarūpa nadīya dvīpoṃ kā nirmāṇa hotā hai.

Strategy for splitting complex sentences:

* Complement clauses will be splitted following the rules stated below-

A. sentential or clausal complement will be an independent sentence.

B. yaha ‘this’ will be added with the clause containing the main verb.

C. yaha ‘this’ will co-refer the entire complement clause.see [here](#_8t3vsrcjgm0) for detail.

Original Sentence

| Sent\_ID\_1 | ***#*** hīrā ne kahā ki ūṁṭa mileṃge.  ‘Hira said that the camel will be available there.’ |
| --- | --- |

After segmentation

| Sent\_ID\_1a | hīrā ne yaha kahā  ‘Hira said this.’ | | |
| --- | --- | --- | --- |
| Sent\_ID\_1b | ūṁṭa mileṃge  ‘Camel will be available there.’ | | |

Complement Clause may occur as following -

Original Sentence

| Sent\_ID\_1 | ***#*** hīrā ne **itanā** kahā ki ūṁṭa mileṃge.  ‘Hira said that the camel will be available there.’ |
| --- | --- |

We adopt the strategy of segmenting such sentences as following

| Sent\_ID\_1a | hīrā ne **itanā** kahā  ‘Hira said this.’ | | |
| --- | --- | --- | --- |
| Sent\_ID\_1b | ūṁṭa mileṃge  ‘Camel will be available there.’ | | |

* **itanā…ki as discourse connective**

itanā…ki may occur as a discourse connective as well. We segment them as following -

Original Sentence

| Sent\_ID\_1 | #nadī ke bāhya taṭa yā natodara taṭa kā itanī tejī se aparadana hotā hai ki visarpa lagabhaga pūrṇa vatta bana jātā hai| |
| --- | --- |

We split such sentences and postulate ‘isase’ as discourse connective in the segmented sentence which brings ‘pariNama’ relation and add iwanA\_ki in the speaker's view row. See [here](#_dar6m74j2o7s) for detailed USR annotation strategy.

After sentence segmentation

| Sent\_ID\_1a | #nadī ke bāhya taṭa yā natodara taṭa kā tejī se aparadana hotā hai |
| --- | --- |
| Sent\_ID\_1b | #isase visarpa lagabhaga pūrṇa vatta bana jātā hai| |

* When two clauses are connected with a connective, we split the sentence into two independent sentences and retain the connective in the sentence where it originally is.

Original Sentence

| Sent\_ID\_1 | ***#*** merī sāikila suṃdara hai lekina abhī vaha gaṃdī hai  ‘My cycle is beautiful but it is dirty now.’ |
| --- | --- |

After sentence segmentation

| Sent\_ID\_1a | ***#*** merī sāikila suṃdara hai  ‘My cycle is beautiful’ |
| --- | --- |
| Sent\_ID\_1b | # lekina abhī vaha gaṃdī hai.  ‘But it is dirty now’ |

Original Sentence

| Sent\_ID\_2 | ***#***rām bīmāra hai isalie vaha skūla nahīṃ gayā  ‘Ram is sick. Therefore he did not go to school.’ |
| --- | --- |

After sentence segmentation

| Sent\_ID\_2a | ***#***rām bīmāra hai ‘Ram is sick’ |
| --- | --- |
| Sent\_ID\_2b | #isalie vaha skūla nahīṃ gayā  ‘He did not go to the school’ |

Original Sentence

| Sent\_ID\_3 | ***#***rāma skūla nahīṃ gayā kyoṃki vaha bīmāra hai  ‘Ram did not go to the school because he is sick. |
| --- | --- |

After sentence segmentation

| Sent\_ID\_3a | ***#***rām skūla nahīṃ gayā ‘Ram did not go to school.’ |
| --- | --- |
| Sent\_ID\_3b | #kyoṃki vaha bīmāra hai ‘Because he is sick.’ |

* When two clauses are connected with a paired connective, we split the sentence into two independent sentences and retain the connective in the main clause.

Original Sentence

| Sent\_ID\_4 | ***#*** yadi āpa mujhe āmaṃtrita karate haiṃ to maiṃ āpake ghara āūṃgā  ‘If you invite me then I will come to your house.’ |
| --- | --- |

After sentence segmentation

| Sent\_ID\_4a | ***#***āpa mujhe āmaṃtrita karate haiṃ ‘You invite me.’ |
| --- | --- |
| Sent\_ID\_4b | #to maiṃ āpake ghara āūṃgā ‘Then I will come to your house’ |

The annotation of discourse connective is presented in the [Discourse Connective Relation](#_86q3x9hey60x) section to ensure no loss of information.

Different Rows of USR

# Row 1: Original Sentence

* All the sentences have a unique ID [**LanguageName\_NameoftheBook\_ChapterID\_SentenceID**] which is followed throughout for maintaining the reference. No space will be given between chapter/ sentence and number. However, language name, book name, chapter and sentence ID will be separated by ‘**\_’.**
* The 1st row is commented with a ‘**#**’ symbol.
* It contains the original sentence in Roman Indic script and in the original script such as Devanagari script for Hindi.

| **Unique sentence ID** | **Original Sentence** |
| --- | --- |
| **Hin\_ABC\_Chapter1\_001**  **[ABC=name of the book]** | **Row 1: *# राम बस अड्डे पर एक पुराने दोस्त के साथ ही बात कर रहा था ।***  **Row 1: *# rāma basa aḍḍe para eka purāne dosta ke sātha hī bāta kara rahā thā.*** |
|

Table 3. Representation of row 1 in USR

### **Sub-sentence Identification**

If the sentence is a title, a section heading or a term combined with its definition, we encode the information in the sentence ID.

**TITLE:** It occurs only once in the discourse, i.e.-the title of the chapter.

The sentence type will be **TITLE**

| Sentence Id | Sentence |
| --- | --- |
| Recipe\_1TITLE | #harī mirca kī caṭanī resipī banāne kī vidhi: |
| Geo\_nios\_7ch\_0079TITLE | #paryatana kenxra: |

**Heading:** All sections and subsections heading are annotated as ‘H’ in the sentence-id.

The sentence-type will be **heading**.

| Sentence Id | Sentence |
| --- | --- |
| Geo\_ncert\_10stnd\_2ch\_0012H | #भारत में वनस्पतिजात और प्राणिजात |

**Term:** If a term is defined, we split the term and its definition into two sub-sentences and specify **‘T’** in the sentence id and sentence type as **Term**.

| Sentence Id | Sentence |
| --- | --- |
| Hin\_Geo\_nios\_7ch\_0029**T** | valita parvata: |
| Hin\_Geo\_nios\_7ch\_0029 | hama pichale pāṭha meṃ paढ़ cuke haiṃ ki pṛthvī kī āntarika halacaloṃ ke kāraṇa paratadāra śailoṃ meṃ valana paDte haiṃ| |

**Fragment:** If a sentence is coming as a fragment, we will specify ‘F’ in sentence ID after the sentence number and declare the sentence-type as ‘fragment’ in sentence-type row.

| Sentence ID | Sentence |
| --- | --- |
| Geo\_nios\_8ch\_0xxx | #vibhinna sāgaroṃ evaṃ mahāsāgaroṃ meṃ lavaṇatā meṃ antara ke mukhya kāraṇa haiṃ |
| Geo\_nios\_8ch\_0xxy**F** | #vāṣpīkaraṇa kī dara| |
| Geo\_nios\_8ch\_0xxy**F** | #nadiyoṃ tathā himakhaṃḍoṃ ke phalasvarūpa tāje jala kī āpūrti| |
| Geo\_nios\_8ch\_0xxz**F** | #mahāsāgarīya jaloṃ kā āpasa meṃ milanā| |

# Row 2: Concepts

Concepts are the semantic constructs. Each entry to the concept row is an unambiguous representation of a concept.

### What to present in the concept row?

* Entity (physical and abstract): laḍakā ‘boy’, pariṇāma ‘result’
* Event : bola ‘speak’, nāca ‘dance’
* Modifier of
  + Entity : acchā ‘good’, thoḍā ‘little’, 10 'ten'
  + Event : dhīre ‘slow’

**Note**

* Spatio-directional terms can have both nominal and relational usages in Indian languages which is discussed [here](#_w5lrt3rmh8th)**.** Nominal usage of these terms are represented in the concept row.
* Negation for now is represented in the concept row.
* If kartā is missing in the original sentence when the sentence is in active mode, kartā will be added in the concept row.

Example, rāma ne eka kelā khāyā aura khelane gayā| ‘Rama ate one banana and went to play. This sentence will be simplified and represented in concept row as following-

| sent\_1a | rāma ne eka kelā khāyā. | | | |
| --- | --- | --- | --- | --- |
| Concept row | rāma | 1 | kelā\_1 | khā\_1-yā\_1 |
| sent\_1b | aura khelane gayā. | | | |
| Concept row | wyax | khela\_1 | jā\_1-yā\_1 |  |

Sent\_1b includes the concept id for ‘vaha’ (which is wyax) even though it is not there in the original sentence.

* Concept ID is not given for named entities, i.e.- proper nouns etc. The root form or bare form of these without concept\_ID is used as a concept. Such as- **himAlaya, rAma** etc.
* Where samanadhikaraNa relation could be found between two concepts, then they will be treated as two different concepts. Such as- himAlaya parvawa will be treated as two different concepts- ***himAlaya* and *parvawa***
* A full reduplicated concept will get only one word as concept and morpho-semantic row will get a ‘dvitva’ tag for generating the reduplication. Such as -

**Gara\_1+Gara\_1** will be represented as **Gara\_1** in concept row and morpho semantic row will get ‘**dwitwa**’ tag.

* Noun compound can come in corpus in three different way, and their representation in concept row will be accordingly -

| **Corpus** | **Concept** | **Example in Corpus** | **Representation in Concept row** |
| --- | --- | --- | --- |
| A B | A\_1+B\_1 | gqha SikRaka | gqha\_1+SikRaka\_1 |
| A-B | A\_1+B\_1 | gqha-SikRaka | gqha\_1+SikRaka\_1 |
| AB | AB\_1 | gqhaSikRaka | gqhaSikRaka |
| AB | A\_1+B\_1 [only A is modified by a modifier not B | varNanawwamaka **BUgolavewwA** | varNanawwamaka\_1+ **BUgola**\_1+**vewwA**\_1 |

* Concept id is not given to pronouns. Pronouns will be marked either by discourse participants, such as-addressee and speaker for 1st and 2nd person pronouns or by root word of the pronouns, such as, *wyax, yax, kim* respectively for 3rd person pronominal, relative pronoun and interrogative pronoun.
* When a concept is a foreign word, we do not transliterate it into wx notation. We use a flag of ‘**^**’ symbol followed by the concept and concept ID. This symbol is used to mark it as a foreign word. Such as - **^word\_1**
* When a concept is an acronym or abbreviation, we transliterate it into wx notation. We use a flag of‘**@**’ symbol followed by the concept and concept ID. This symbol is used to mark it as an acronym/ abbreviation. Such as - **@eic.sl.yU., @nAsA**
* We do not represent the concept of symbols as a concept, we unfold their semantics. Such as - ₹ 500 will be represented in concept as **rupayA\_1**+500, 78% will be 78+**prawisawa**\_1
* All numbers would be represented as digits when they represent cardinal relation. Such as - **1** Ama, **2** seba.
* When *eka* is used for one, it will be represented as ‘**1**’, such as -

rAma ne eka seba KAyA

rAma, **1**, seba\_1, KA\_1-yA\_1

However, when *eka* is used as an indefinite article ‘a/ an’, then it is represented as ‘**eka\_2’**, such as-

rAma eka xoswa se milA

rAma, **eka\_2**, xoswa\_1, mila\_1-yA\_1

### 

### How are concepts represented?

The format : Root\_Concept\_ID

#### Why Concept ID?

Concepts could be ambiguous in nature. For example, the lexeme *paḍha* expresses two concepts: 'study’ (as in *The boy studies in 7th standard*) and ‘read’ (‘*the boy reads a book*’). To resolve this kind of ambiguity at the conceptual level, every concept gets a unique concept ID in the concept dictionary. Each unique concept is mapped with an appropriate equivalent from other languages. They are listed in the ***concept dictionary***. For entries in the concept dictionary, see [here](#_h4olo5nsx827).

Concept dictionary contains concepts in wx-format which is mapped with Roman Indic script [here](#_4jb4zeh1q42w).

| Original Sentence | ***rāma basa aḍḍe para eka purāne dosta ke sātha hī bāta*  *nahīṃ*** ***kara rahā thā*** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Concept row | **rāma** | **basa\_1+aḍḍā\_1** | **eka\_2** | **purānā\_1** | **dosta\_1** | **nahīṃ\_1** | **bāta+kara\_1-0\_rahā\_thā\_1** |

Table 4. Representation of row 2 in USR

Appropriate concepts are to be fetched from the concept dictionary and specified in the USR.

| What to give concept ID | What not to give Concept ID |
| --- | --- |
| Common noun | pronominal |
| Compound Noun | Numex as cardinal number |
| verb |  |
| Modifier of verb and noun | Reflexive and reciprocal pronoun |
|  |  |
| Apasa |  |
| Measurement unit  Such as-   * 80+kimi**\_1** * 2+GaMtA**\_1** |  |
| Symbols, Such as - ;₹’ - **rUpiyA\_1** |  |
| eka\_2  When ‘eka’ is an indefinite |  |

How do we represent Multi Word Expression in Concept row-

| MWE type | Example |
| --- | --- |
| Noun Compound  (composition of   1. two/more nominal concepts (and they do not together form a named entity) 2. NC with two modifier and one head 3. NC with modifier of modifier   OR   1. named entity and one/more nominal concepts | **pramANa\_1+pawra\_1**  **skula\_1+sikRaka\_1+samiwi\_1**  **vanya\_1+jIva\_1+jAwi\_1**  **banArasa+hindu+viSvavixyAlaya+ CAwrAvAsa\_1**  **kolakAwA+police\_1**  **kolakAwA+police\_1+samuxAya\_1** |
| Named Entity | **banArasa+hindu+viSvavixyAlaya** |
| Partial Reduplication | **xina+prawixina\_1** |
| Echo-word | **Asa+pAsa\_1** |
| Adj+particle | **CotA\_1+sA\_1** |
| Noun+morpheme together making adjective | **namaka\_1+yukwa\_1** |
| Frozen expression | **cAroM+Ora\_4**  **eka+sAWa\_1**  **kriyA+kalApa\_1**  **eka+jEsA\_1** |
| Phrasal expression | **mote+wora+para\_1** |
| With particle | **hAla+hI\_1** |
| Measurement expression | **4+kilo\_1**  **5+GaMtA\_1**  **4+digrI+selasiyAsa\_1**  **ciwra\_1+2.1** |
|  |  |

##### **Note**

* ciwra saMKyA 2.1 will be presented as two concepts as following-

ciwra\_1+saMKyA\_1 as one concept and 2.1 as another concept with ‘rs’ relation

* *bahuwa sA/s*e will be ***bahuwa sArA***

#### Which items are not presented in the concept row

| Name of Category | Example |
| --- | --- |
| Connectives | aura ‘and’, yā ‘or’, kyoṃki ‘because’, isaliye ‘therefore’ |
| Discourse Particle | hī, bhī, sirpha |
| Comparative and superlative marker | * tara, - tama |
| Comparative and superlative word | sabse, adhika, jyādā, kama |
| vālā | When vālā comes as suffix and makes the concept adjective |
| Post-positions | ne, ko, para, … |
| Light verb in V-V compound Verb or Raṃjaka kriyā | khā **le** ‘ate’  sajā **de** ‘decorate’ |
| Salutation or Respect marker | jī, srī, Mr., Ms. … |
| kim in yn\_interrogative sentence | **kyā** āpane cāvala khāye? ‘Did you eat rice?’ |

Table 5. What not to represent in Concept row, i.e. row 2

We will discuss below the representation of different kinds of concepts.

#### Entity

An entity can be a simple item or it can be a composite idea. For example, kitāba ‘book’ refers to some singular item in the real world whereas ‘textbook’ is a composite idea which refers to “a book that contains detailed information about a subject for people who are studying that subject” [Cambridge Dictionary]. ‘History textbook’ is more composite in nature that refers to a textbook about History. ‘Ancient History textbook’ further specifies the period of history.

A composite idea can be expressed as a multi-word expression (MWE) in many languages. One such MWE is Noun Compound (NC).

An NC is made up of a head noun and one or more noun modifiers. In USR, they are joined with "+".

basa aḍḍā -> basa\_1+aḍḍā\_1

If an NC occurs with - (hyphen) then give the concept ID at the end of the NC e.g. bhū-kṣetraphala\_1. See [Appendix-3](#_ljybm050mn7q) for further details.

.

### Pronominal Reference to an Entity

Resolution of pronominal expression (i.e. determining which entity it refers to) happens at the discourse level through co-referencing (For further discussion see [Pronominal co-reference](#_5bcv7oxrdwgu)) and deictic pronouns through annotating deicticity in the [Speaker’s view](#_5d12jaw9hs6j) row. In the concept row, following label is annotated:

### Personal Pronoun

**1st person**- ‘speaker’

**2nd person**- ‘addressee’

##### How to distinguish among tuma, tū and āpa (the three forms of 2nd person pronoun)

In concept row, all three forms of 2nd person pronoun are marked as addressee. The distinction is encoded in the Speaker's view row under [Respect and Informal information Section](#_hua3hey2sm50).

###### wyax

3rd Person pronominals are conceptualized as *wyax* in concept row. *wyax* represents all such variables which have the form of any third person pronoun and either have trace in the text or in the discourse.

##### How to distinguish between proximal and distal form of wyax

wyax can have two separate forms, proximal and distal (yaha and vaha) and in some languages, there is another additional form of wyax, that is far-distal. The distinction is encoded in the Speaker's view row under [Proximal and Distal Section.](#_qc3blhhzii6f)

###### Reciprocal Pronouns

Reciprocal pronouns, which express mutual relationship, would be treated as a compound word. Such as- each-other. In Hindi, *eka dūsare* is an example of a reciprocal pronoun.

###### Reflexive Pronoun

Reflexive pronouns are pronouns which refer back to the nominal itself, such as- myself, himself, herself etc. In Hindi, *khuda, apanā* are examples of reflexive pronouns.

###### Interrogative Pronoun

All interrogative pronouns are conceptualized as *kim* in concept row. See [here](#_r2chpsyqtke1) for Hindi interrogative pronoun list.

###### Relative Pronoun

All relative pronouns are represented as *yax* in the concept row. However, relative pronouns which function as a discourse connective are not represented in the concept row.

##### Example for all Pronouns

| Personal Pronoun | 1st Person | *# तुम* ***मेरे*** *घर आओ. # tuma mere ghara āo* | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | addressee | **speaker** | | | ghara\_1 | | | ā\_1-o\_1 |
| 2nd Person | *#* ***तुम*** *मेरे घर आओ. # tuma mere ghara āo* | | | | | | | |
| Concept | **addressee** | speaker | | | ghara\_1 | | | ā\_1-o\_1 |
| 3rd Person Pronominal | | #**उसने** नहीं खाया । #usane nahīṃ khāyā. | | | | | | | |
| Concept | | **wyax** | nahīṃ\_1 | | | | | | khā\_1-yā\_1 |
| Reciprocal Pronoun | | *#हम एक दूसरे से प्यार करते हैं ।*  *#hama eka dūsare se pyāra karate haiṃ.* | | | | | | | |
| Concept | | speaker | **eka+dūsarā** | | | | | | pyāra+kara\_1-tā\_hā\_1 |
| Reflexive Pronoun | | #unhoṃne apanā kāma khuda kiyā | | | | | | | |
| Concept | | wyax | **apanā** | | kāma | | **khuda** | | kara\_1-yā\_1 |
| Interrogative Pronoun | | #राम क्या खा रहा है? #rāma kyā khā rahā hai? | | | | | | | |
| Concept | | rāma | **kim** | | | | | | khā\_1-0\_rahā\_hai\_1 |
| Relative Pronoun | | #भारत का सबसे दक्षिणी बिंदु, जो इंदिरा बिंदु कहा जाता था, सन्‌ 2004 में जलमग्न हो गया ।  ‘The southernmost point of India, which was known as Indira point, was submerged in the water in the year 2004. | | | | | | | |
| Concept | | bhārata | dakṣiṇī\_1 | biṃdu\_1 | yax | iṃdirā+biṃdu\_1 | kaha\_1-yā\_jātā\_thā\_1 | san+2004 | jalamagna+ho\_1-gayā\_1 |

Table 6. Representation of different pronominal concepts

###### Events

##### Concepts for verb forms consist of root and TAM info separated by ‘-’. `

* Verbs can be stative or action verbs.
* A complex predicate consists of a kriyāmūla and a kriyā They are joined by "+" and the whole predicate consists of kriyāmūla and kriyā gets a concept label after the kriyā.
* Non-finite verbs are represented in root form without TAM specified in the concept row

Example

| **Stative verb** | #राम अच्छा **है** । #rāma acchā hai. | | | | |
| --- | --- | --- | --- | --- | --- |
| Concept | rāma | ācchā\_1 | **hai\_1-pres** | | |
| **Action verb** | *#राम खीर* ***खा रहा है*** *। #*rāma khīra khā rahā hai. | | | | |
| Concept | rāma | khīra\_1 | **khā\_1-0\_rahā\_hai\_1** | | |
| **Complex Predicate** | #राम ने नदी मे **स्नान किया** । #rāma ne nadī me snāna kiyā. | | | | |
| Concept | rāma | nadī\_1 | **snāna+kara\_1-yā\_1** | | |
| **Non-finite verb** | #राम ने स्कूल **जा कर** खाना खाया #rāma ne skūla jā kara khānā khāyā | | | | |
| Concept | rāma | skūla\_1 | **j**ā**\_1** | khānā\_1 | khā\_1-yā\_1 |
| **Non-finite verb** | #गायों के **दुहने** के लिये राम घर गया। #gāyoṃ ke duhane ke liye rāma ghara gayā | | | | |
| Concept | gāya\_1 | duha\_1 | rāma | ghara\_1 | ja\_1-yā\_1 |

Table 7. Representation of different eventualities

### Tense, Aspect, Modality (TAM)

Like content words, Tense-Aspect-Mood markers (henceforth, TAM string) can also be polysemous in nature. That is why they are also represented in the TAM concept dictionary with unique ids (See [Appendix-5](#_wya5z716yvfp)). A verb can be of tinganta (तिङन्त/finite) or kridanta (कृदन्त/non-finite, infinitival, verbal noun, participial) form.

**Note**

* The TAM string is separated from the root by '-' (hyphen) when it is a finite verb.

Example-*kara\_1****-gā\_1***→ karegā

* The multiword TAM string is written with an underscore.

Example- *kara\_1****-yā\_thā\_1***→ kiyā thā

* The default form of TAM occurs in 3rd person singular form.

Example- *kara\_1****-gā\_1***→ karegā/ karegī/ karoge

*kara\_1****-yā\_thā\_1***→ kiyā thā/ kiyi thī/ kiye the

* When a bare form of a verb is followed by a TAM marker, we postulate a zero in the initial slot of the TAM string.

Example. ***kara\_1-0\_rahā\_hai\_1*** →kara rahā hai/ kara rahī hE/ kara rahī ho

* For imperative sentences, TAM will be by default, ‘o’.

tū ghara **jā**,

addressee, ***jā\_1-o\_1***.

* For the correct Hindi TAM see [here](#_wya5z716yvfp).

##### **Modifiers of Entities**

Adjectives or the modifier of entities could be an adjective, a quantifier or a cardinal/ ordinal number or an intensifier. They are represented in bare form and get concept id.

Some relations are listed below:

| **Semantic Role** | **Tag** | **Example** |
| --- | --- | --- |
|  |  |  |
| Modifier | mod | purānā ‘old’, motā ‘fat’, sapheda ‘white’, |
| Quantifier | quant | pratyeka ‘every’, kucha ‘some’, eka ‘some’ |
| Cardinal number | card | 1 ‘one’, 2 ‘two’ |
| Ordinal number | ord | pahalā ‘first’, dūsarā ‘second’ |
| Intensifier | intf | bahuta ‘very’ |

Table 8. Modifier of entities and their tags

Examples of modifier of entities

| Adjective | #राम एक पुराने दोस्त से बात कर रहा था।  #rāma purāne dosta ke sātha bāta kara rahā thā. | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept row | rāma | eka\_2 | **purānā\_1** | dosta\_1 | | | bāta +kara\_1-0\_rahā\_thā\_1 | | |
| Quantifier | #सब लड़के आएंगे #saba laḍake āeṃge | | | | | | | | |
| Concept row | **saba\_1** |  | laḍakā\_1 | ā\_1-gā\_1 | | | | | |
| Intensifier | #बहुत मोटी बिल्ली दीवार पर सो रही है. #bahuta moṭī billī dīvāra para so rahī hai. | | | | | | | | |
| Concept row | **bahuta**\_1 | | | motā\_1 | | bīllī\_1 | | dīvāra\_1 | so\_1-0\_rahā\_hai\_1 |
| Cardinal Number | #राम रोज़ दो सेब खाता है ।#rāma roja do seba khātā hai. | | | | | | | | |
| Concept | rāma |  | roja\_3 | **2** | | seba\_1 | | khā\_1-tā\_hai\_1 | |
| Ordinal Number | #राम दशरथ के प्रथम पुत्र हैं. #rāma daśaratha ke prathama putra haiṃ | | | | | | | | |
| Concept | rāma |  | daśaratha | **prathama\_1** | | putra\_1 | | hai\_1-pres | |

Table 9. Representation of modifier of entities in USR

##### **Modifiers of Events**

Adverbials are the modifier of events which provide information on the manner of adverbs, negations etc. We represent manner adverb and negation in concept row with appropriate concept ID.

Examples of modifiers of events

| kriyā viśeṣaṇa (manner adverb) | #राम भागकर आया #rāma bhāgakara āyā | | |
| --- | --- | --- | --- |
| Concept | rāma | **bhāga\_1** | ā\_1-yā\_1 |
| Negation | #राम नहीं आएंगे #rāma nahīṃ āeṃge | | |
| Concept | rāma | **nahīṃ\_1** | ā\_1-gā\_1 |

Table 10. Representation of modifier of events in USR

**Note:**

Check the Concept dictionary for the correct concept label ID.

# Row 3: Index for the Concepts

This row in USR gives an indexing where each concept (i.e. the prakrti) is indexed according to the place of occurrences, represented in the concept row. This indexing helps to mark the head-dependency, co-referencing and compositionality among members of concepts which we will discuss in the Dependency row, Discourse elements row and Construction row.

| Original Sentence | #राम बस अड्डे पर अपने एक पुराने दोस्त के साथ ही बात कर रहा था ।  #rāma basa aḍḍe para apane eka purāne dosta ke sātha bāta kara rahā thā. | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | rāma | basa\_1+aḍḍā\_1 | apanā | eka**\_2** | purānā\_1 | dosta\_1 | bāta+kara\_1-0\_rahā\_thai\_1 |
| **Index** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| Original Sentence | #प्रशांत महासागर सबसे बड़ा महासागर है। #praśanta mahāsāgara sabase baḍā mahāsāgara hai. | | | | | | |
| Concept | praSānta\_1+mahāsāgara\_1 | baḍā\_1 | mahāsāgara\_1 |  | hai\_1-pres | | |
| **Index** | **1** | **2** | **3** |  | **4** | | |

Table 11. Representation of Index row, i.e.-row 3

# Row 4: Semantic Categories of Nouns

The Semantic category row specifies the semantic category of a concept.

* Currently, four generic named entity categories are being annotated, namely- *per*(son), *org*(anisation), *place* and *ne*. *ne* is the underspecified tag used for all such named entities which do not fall into the category of either person, place or organization. Apart from that, we mark *Time, number,* and *animacy* categories.
* This row also captures the gender information which is an ontological information. Only inherent gender is marked and grammatical gender is not marked in USR.
* For speaker and addressee if in singular number, gender will be marked as per the context.
* For speaker and addressee, if in plural number, gender will be not specified. Such as, baccA ‘children’ will get only ‘anim’ information and no gender information will be annotated for them.
* *wyax* [3rd person pronominal form] will not get animacy or gender information as the information will be mapped from co-reference.

|  | **Semantic Category** | **Tag** | **Example** |
| --- | --- | --- | --- |
| Named Entity | Person name, a subset of animacy | per/male  per/female | rāma ‘Rama’, karabi ‘Karabi’ |
| Place (City, Continent) name | place | dillī ‘Delhi’ |
| Organization name | org | banārasa hindū yūnivarsiṭī 'Banaras Hindu University’ |
| Names of movies, medicine, cuisine, games, disease | ne | phauṭabala ‘football’ |
| Foreign Word | Foreign words | fw | forest\_1+principle\_1 ‘Forest Principle’ |
| Time entity | day\_of\_week | dow | Śukravāra ‘Friday’ |
| month\_of\_year | moy | agasta ‘August’ |
| year\_of\_century | yoc | 1947, san\_1+2004, |
| century | era | 17+saxI\_1, SawAbxI ISA pUrva |
| date\_of\_month | dom | 15th |
| calendric\_unit | calendricunit | 11+ tārīkha\_1 ‘11th date’ |
| clock\_time | clocktime | 5+baje\_1 ‘5 o’clock’ |
| Season of a year | season | SIta\_1 ‘winter’, basanta\_1 ‘spring’ |
| Any special day | timex | Independence Day, Christmas Day |
| Number entity | measurement | meas | 5+kilo\_1, 10+meter\_1, 10+GaMtA\_1 |
| count | numex | 2, laḍakA\_1 ‘two boys’ |
| animacy | living beings unless a proper noun | anim | Speaker, addressee, laḍakā ‘boy’, |
| Gender | Gender information of living being with inherent gender | male  female | **sītā** baccoṃ ko phala detī hai ‘Sita gives fruit to the children.’ |

Table 12. Contents of Semantic Category Row

#### Example of Semantic Category Row

| Original sentence | *#*अर्जुन बनारस के हिंदू विश्वविद्यालय में 10th अगस्त, 2021, शुक्रवार शाम को 5 बजे अध्यापक के रूप में नियुक्त हुए.  *#*arjuna banārasa ke hiṃdū viśvavidyālaya meṃ 10 agasta, 2021, śukravāra śāma ko 5 baje adhyāpaka ke rūpa meṃ niyukta hue | | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| concept | arjuna | | | banārasa | | | hiṃdū+ viśvavidyālaya | 10 | | | agasta | | | 2021 | | | śukravāra | | | śāma\_1 | 5+baje\_1 | adhyāpaka\_1 | niyukta+ho\_1-yā\_1 |
| Semantic Category of Noun | **per male** | | | **place** | | | **org** | **dom** | | | **moy** | | | **yoc** | | | **dow** | | |  | **clocktime** | **anim** |  |

Table 13. Representation of Semantic Category Row

# 

# Row 5: Morpho-Semantic Information

At the Morpho-Semantic row, the speaker’s *vivakṣā* (intention) to compare, causativize and adjectivize are encoded which, during language generation, are mostly represented in terms of a derived form of the root word that denotes the given concept.

| Number | pl-plural | rāma kala kaī **chātroṃ[pl]** se mile  ‘Ram met many students yesterday.’ |
| --- | --- | --- |
| mawup | Modifier derives from the root with affixation, such as -valā | pūrṇa **caṃdramā vālī** rāta ko pūrṇimā kahā jātā hai  ‘The night of the full moon is called Purnima.’ |
| kqw | Predicative past perfective modifier, occurs on predicate position and modifies the *kartā* | paraṃtu ye aṃtarnirbharatāoṃ ke jaṭila jāla dvārā eka taṃtra meṃ **guṁthī** huī haiṃ  ‘But, this is closely integrated in a system through multiple networks of interdependencies.’ |
| compermore | Comparative degree marker | gaṃgā yamunā se **jyādā** laṃbī hai  ‘Ganga is **longer** than Yamuna’ |
| comperless | Comparative degree marker | rāma mohana se **kama** buddhimāna hai.  ‘Ram is **less** intelligent than Mohan.’ |
| superl | Superlative degree marker | gaṃgā bhārata kī s**abase baḍī** nadī hai  Ganga is the **largest** river in India |
| dvitva | Full reduplication | **Gara Ga ra** me citTi Ayi  ‘Letter reached every house.’ |
| causative | Morphological causativization | māṃ ne bacce ko khānā **khilāyā**.  ‘The mother **fed** the baby.’ |
| doublecausative | Morphological double causativization | māṃ ne rāma se bacce ko khānā **khilavāyā**  ‘The mother **fed** the baby by Rama’. |

* Table 14. Contents of Morpho-Semantic Category Row

#### 

#### General Information

* We only mark the plurality information by ‘pl’.
* Numbers are marked only for countable nouns. For all other kinds of nouns, numbers information can be left blank.
* For 1st and 2nd person pronouns, i.e.-speaker and addressee, we give number information to distinguish between singular and plural.
* For *wyax*, or 3rd person pronominal, we do not give number information as information will be mapped from co-reference.

|  | *#*आपने और मैंने राम को ज्यादा लंबे और सबसे महंगेवाले चावल दिये.  *#*āpane aura maiṃne rāma ko **jyādā** laṃbe aura sabase mahaṃge**vāle** cāvala diye | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept row | addressee | | | | | | speaker | | | | | | rāma | | | | | | laṃbā\_1 | mahaṁgā\_1 | cāvala\_1 | | | | | | de\_1-yā\_1 | | | | | |
| Morpho-Semantic Information |  | | | | | |  | | | | | |  | | | | | | **compermore** | **mawup** |  | | | | | |  | | | | | |
|  | *#*एक गाँव में तीन लड़के रहते हैं *#*eka gāṁva meṃ tīna laḍake rahate haiṃ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Concept | eka\_2 | | | | | | | | gāṁva\_1 | | | | | | | | 3 | | | | laḍaka\_1 | | | | raha\_1-tā\_hai\_1 | | | | | | | |
| Morpho-semantic information |  | | | | | | | |  | | | | | | | |  | | | | pl | | | |  | | | | | | | |

* Table 15. Representation of Morpho-Semantic Category Row

|  | *#*māṃ ne bacce ko khānā **khilāyā**. | | | |
| --- | --- | --- | --- | --- |
| Concept row | māṃ\_1 | baccā\_1 | khānā\_1 | **khā\_1-yā\_1** |
| Morpho-semantic information |  |  |  | **causative** |

* Table 16. Representation of Causative verb

# Row 6: Dependency Relation

# In this row we mark the relation between head and dependent in following way-

Index of the head: the relation of dependent with the head

Here, the relation means "what the dependent is to the head". So, if we get a relation as 2:k2, this means '2' is the index of head and 'karma' or 'k2' is the name of the relation of the dependent with the head or dependent is k2 of head.

Two types of head-dependency relations are captured in this row. They are

* **kāraka relation** between verbs and its dependent nouns
* **kāraketara** (Other than kāraka) relations between
  + verb and its other non-kāraka dependents
  + Noun and its modifiers

**Notes**

* The tags for kāraka relations start with ‘k’
* The tags for Other than kāraka relations start with ‘r’
* The head or *mukhya viśeṣya* in the dependency tree is marked as **0:main.** Generally it is realized as the finite verb in a sentence. However, in case of fragment, title, term- the *mukhya viśeṣya* may not be a finite verb but still there will be head-dependency relation and head will get **0:main**
* The relation between the *viśeṣya* or head and its dependents are specified in this row as the index **of the head : relation tag** in the column of the dependent**.**
* The convention followed in this document is to mark the **dependen**t in **bold** and underline the head.

## kāraka relations

* All kāraka relations start with ‘k’ and are followed by a numerical

A list of six main kāraka relations given below

| **kāraka** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| *kartā* | k1 | most independent participant of an action | **rāma** āma khātā hai  ‘Ram eats mango.’ |
| *karma* | k2 | locus of the result of the action | mōhana nē **āma** kharīde  ‘Mohana bought mangoes.’ |
| *karaṇa* | k3 | The instrument required for the performance of the action | ratanā ne āma **cākū se** kāṭe  ‘Ratna cut the mangoes with a knife.’ |
| *saṃpradāna* | k4 | recipient/beneficiary | **billī ko** dūdha do  ‘Give milk to the cat.’ |
| *apādāna* | k5 | Source | **peḍa se** eka pattā girā  ‘A leaf fell from the tree.’ |
| *Viṣayaa*  *dhikaraṇa* | k7 | Location elsewhere | ve **rājanīti** **para** carcā kara rahe the  ‘They were discussing politics.’  maiṃ **rāma ke bāre meṃ** nahīṃ jānatā  ‘I do not know about Ram.’ |
| *kāladhikaraṇa* | k7t | Time of the event | rāma **cāra baje** āyegā  ‘Ram will come at 4 o’clock.’ |
| *deśadhikaraṇa* | k7p | Locus of the event | **meja** para kitāba hai  ‘The book is on the table.’ |

Table 16. six main kāraka

Note - Although, karta/ karma of Complex Predicate gets a genitive relation, but the dependency relation is k1/k2. See Complex Predicate for details

### exemplifyOther kāraka relations

| **Relation Name** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| *anubhava-kartā* | k4a | Experiencer | **rāma ko** āma pasaṃda hai  ‘Ram loves mango.’ |
| *g*a*uṇa karma* | k2g | Secondary object | ve **gāṃdhījī ko** bāpū kahate haiṃ  ‘They call Gandhiji Bapu.’ |
| *destination* | k2p | Destination or goal | rāma **ghara** gayā  ‘Ram went home.’ |
| *prakṛti apādāna* | k5prk | Source material | jūte **camaḍe** **se** banate haiṃ  ‘Shoes are made of leather.’ |
| *prayojaka kartā* | pk1 | causer | **māṁ ne** bacce ko khānā khilāyā  Mother fed the babies.’ |
| *prayojya kartā* | jk1 | causee | māṁ ne āyā se **bacce** **ko** khānā khilavāyā  ‘Mother made the maid to feed the babies.’ |
| *madhyastha-kartā* | mk1 | mediator causer | māṁ ne **āyā se** bacce ko khānā khilāvāyā  ‘Mother made the maid to feed the babies.’ |
|  | k7a | according to someone/ something | **rāma ke anusāra** sītā ghara para nahīṃ hai  ‘According to Ram Sita is not at home.’ |

Table 17. Other kāraka relations

### Relations associated with kāraka

| *saha-kāraka* | rask1 | *associate of*  *kartā* | **rāma ke sātha** mohana bājāra gayā.  ‘Mohana along with Ram went to the market.’ |
| --- | --- | --- | --- |
| rask2 | *associate of karma* | rāma ne **dūdha ke sātha** kelā khāyā.  ‘Ram ate bananas with milk.’ |
| rask3 | *associate of karaṇa* | vaha **cammaca ke sātha** kāṁṭe se sabjī khā rahā hai  ‘He is eating vegetables with a fork along with a spoon.’ |
| rask4 | *associate of saṃpradāna* | vaha **guru jī ke sātha** śiṣyoṃ ko dakṣiṇā detā hai  ‘Along with the honorable Guru, he gives donations to the disciples.’ |
| rask5 | *associate of apādāna* | **bālakanī** **ke sātha** khiḍakiyoṃ se bhī dhūla ā rahī thī  ‘The dust came from the windows along with the balcony.’ |
| rask7 | *associate of adhikaraṇa* | unhoṃne **rājanītika muddoṃ sahita** anya viṣayoṃ para kitābeṃ likhī haiṃ  ‘He has written books on other topics including political issues.’ |
| *kartā* *samanadhikarana* | k1s | kartā and its viśeṣaṇa resides in the same locus, when the verb is copulative | rāma **buddhimāna** hai  ‘Ram is intelligent.’ |
| *karma samanadhikarana* | k2s | karma and its viśeṣaṇa resides in the same locus | rāma mohana ko **buddhimāna** samajhatā hai  ‘Ram considers Mohan to be intelligent.’ |

Table 18.Relations associated with kāraka

## kāraketara relation

Apart from karaka relations, the head of the sentence can have some non-kāraka relations with its dependents. They are further divided into different categories according to the specification of the relations as discussed below-

### samānādhikaraṇa

| **Relation Name** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| samānādhikaraṇa/ viśeṣaṇa | mod | Modifier or the head and its modifier share the same locus | **moṭī** billī meja para sotī hai  ‘‘The fat cat sleeps on the table’ |
| bhūtakālika  samānādhikaraṇa | rbks | equal locus of the action and the dependent action denoted by non-finite verb) in past tense | maiṃne mohana ke dvārā **likhī huī** kitāba paḍhī  ‘I read the book written by Mohana.’ |
| vartamānakālika samānādhikaraṇa | rvks | equal locus of the action and the dependent action denoted by non-finite verb) in present tense | maine jangal meṃ eka **bhāgate hue** śera ko dekhā  ‘I saw a running lion in the jungle.’ |

Table 19. samānādhikaraṇa

### bhāvalakṣaṇa:

The nominal form of the dependent verb (VN) plays the role of a referent with respect to which the time of the main event (VM) is specified.

| **Relation Name** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| bhāvalakṣaṇa samānakālika | rblsk | Temporal overlapping or co-temporality between VN and VM | rāma ke vana **jāne ke samaya** sītā unakā anusaraṇa karatī hai  **‘**Sita followed Ram while he was going to the forest.’ |
| bhāvalakṣaṇa pūrvakālika | rblpk | VN precedes VM | sūrya **ugane ke bāda** khānā khāo  ‘Eat after the sun rises.’ |
| bhāvalakṣaṇa anantarakālika | rblak | VN follows VM | sūrya **ugane se pahale** nahāo  ‘Bathe before the sun sets.’ |

Table 20. bhāvalakṣaṇa

### kālavācī

| **Relation Name** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| pūrvakālika | rpk | One event occurs after the previous event is done and the kartā is shared | rāma ne khānā **khākara** pānī piyā  ‘Ram drank water after eating a meal.’ |
| samānakālika | rsk | Two events occur simultaneously and the kartā is shared. | rāma **sote hue** kharrāṭe bharatā hai  ‘Ram snores while sleeping.’ |

Table 21. kālavācī

### Spatio-temporal Information

| **Relation Name** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| deśalakṣaṇa | rdl | A space is referent of another locus | **peḍa ke** ūpara cāṁda hai  ‘The moon is above the tree.’ |
| kālalakṣaṇa | rkl | A time is referent of actual temporal information of the event | **7 se** pahalerāma ghara āyā  ‘Ram came home before 7 o’clock.’ |

Table 22. Spatio-temporal information

### Intra-sentential sangati

| **Relation Name** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| *tādarthya* | rt | Purpose of the event | **mohana ke lie** seva lāo  Bring apples for Mohan.’ |
| *kāraṇa* or *hetu* | rh | Reason of an action | **mohana ke kāraṇa** mujhe dera ho gayī  ‘I became late because of Mohan.’ |
| *udāharaṇam* | re | Elaboration or example of an expression | kucha vastuoṃ kā nirmāṇa prakṛti ne kiyā hai jaise **parvata nadiyāṁ prāṇī**  ‘Some things are made by nature like rivers, trees and animals.’ |
| *samānādhikaraṇa* | rs | When one entity equates with the other entity | #pṛthvī kī āntarika paratoṃ kā vargīkaraṇa aura unakī moṭāiyoṃ ko **citra saṃkhyā** 2.1 meṃ darśāyā gayā hai | |

Table 23. Intra-sentential sangati

### Genitive or Possessive relation between two entities

| **Relation Name** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| saṣṭhī | r6 | Genitive | **rāma kā** kitāba  Ram’s book. |
| sthāyī svāmī | rsm | Possessor of some entity | **rāma ke** pāsa kitāba hai**.**  ‘Ram has the book.’ |
| asthāyī svāmī | rsma | Temporary possessor of some entity | **rām**a **ke pāsa** sītā kī kitāba hai.  ‘Ram has Sita’s book.’ |
| Human to human | rhh | Relation between two human beings when there is a stative verb. | **rāma** **ke** do beṭe haiṃ  ‘Ram has two sons.’ |

Table 24. Genitive or Possessive relation between two entities

### Sādrisya, vibhājana and nirdhāraṇa

| **Relation Name** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| sādrisya | ru | When there is comparison between two entities based on the resemblance or similarity | **gulāba** **jaise** phūla pānī meṃ nahīṃ ugate haiṃ  ‘Rose-like flowers do not bloom in water.’ |
| vibhājana | rv | When two entities are compared and there are inequalities observed between them | rādhā **mīrā kī tulanā meṃ** adhika suṃdara hai  ‘Radha is more beautiful than Mira.’ |
| nirdhāraṇa | rn | *‘*nirdhāraṇam or specification is made by separating one from the many by reason of its genus, quality and action*’*. | **gāyoṃ meṃ** kālī gāi sabase jyādā dūdha detī he. ‘Among cows, black cows give the most milk.’ |

Table 25. Sādrisya, vibhājana and nirdhāraṇa

### Other kārakatera relation

| **Relation Name** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| Direction | rd | Direction towards a goal | sītā **gāṃva** **kī ora** jā rahī thī  ‘Sita is going towards the village.’ |
| kriyā viśeṣaṇa | krvn | Manner adverb | rāma **bhāgakara** āyā  ‘Ram came running.’  rāma **dhīre** chalatā hai  ‘Ram walks **slowly**.’ |
| Negation | neg | Negation | rāma **nahīṃ** āyā hai  ‘Ram has not come.’ |
| vakya viśeṣaṇa | vkvn | Sentential adverb | rāma **sāyada** nahīṃ āyā hai  ‘Ram probably has not come.’ |
| frequency | freq | A temporal and manner information of an event which reoccurs over a period of time | vaha **roja** yahāṃ ātā hai  ‘He comes here everyday.’ |
| Negation in Associatives | rasneg | When there is absence of some participant | **jala ke binā** koī bhī jīva jīvita nahīṃ raha sakatā|  ‘No animal can be alive without water.’ |
| Relation path | rp | for "through" or "via" which indicates a path of movement. | karka rekhā isa **mahādvīpa se** **hokara** gujaratī hai ।  ‘The Tropic of Cancer passes through this continent.’ |
| Construction Part | cxnpart | When element/s of a construction shows the relation with other element/s and not defined as a dependency relation due to non- compositionality | rāma do bāra hara **12 ghaṃṭe meṃ** khātā hai  ‘Ram eats twice in every 12 hours.’ |
| Relation address | rad | When there is an address by some name | nāraka! mere īśvara, lepacāoṃ kī duniyā meṃ āpa saṃgīta ke janaka haiṃ  ‘Naraka, my lord, you are the father of songs in the world of Lepchas.’ |

Table 26. Other kārakatera relation

**Note**

In the context of the relation path, the term "relation path tag" will be used there to signify "via,"or "through" or in other words, "से होकर" in Hindi. The sentence is constructed in a manner that conveys this meaning. Here, "से होकर" is taken as a post position. It is important to note that, in this usage, "से होकर," which is typically a verb, is employed in particular semantics as a post position.

### Modifier and modified relations

| **Relations** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| Demonstrative | dem | Point to a specific entity | **yaha** kitāba lāla hai.  ‘This book is red.’ |
| Cardinal number | card | Cardinal numbers or counting numbers | rāma roja **do** seba khātā hai.  ‘Ram eats two apples everyday.’ |
| Ordinal number | ord | Number which represents the position or rank | rāma daśaratha ke **prathama** putra haiṃ  ‘Ram is the first son of Dasaratha.’ |
| Quantifier | quant | A limiting noun modifier express quantity | **saba** laḍake āeṃge  ‘Every boy will come.’ |
| Intensifier | intf | Intensifying quality or quantity of an entity | **bahuta** moṭī billī dīvāra para so rahī hai  ‘The very fat cat is sleeping.’ |
| Quantity more than a certain number | quantmore | when quantity is mentioned not as a specific number but as more than a certain number, then we will not use cardinal relation but quantmore relation | #pṛthvī para **tīna hajāra se adhi**ka vibhinna khanija haiṃ।  There are over three thousand different minerals on this earth. |
| Quantity less than a certain number | quantless | when quantity is mentioned not as a specific number but as less than a certain number, then we will not use cardinal relation but quantless relation | pṛthvī para **tīna hajāra se kama** vibhinna khanija haiṃ  There are less than three thousand different minerals on this earth. |

Table 27. Modifier and modified relations

### Different Measurement relations

| **Relations** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| Duration | **dur** | Measuring the time span, during which something continues | rāma **10 ghaṃṭe** calā  ‘Ram walked for ten hours.’ |
| Extent | **extent** | Measuring the area covered by someone/ something | rāma **10 kimi** calā  ‘Ram walked ten km.’ |
| Quantity | **quant** | Measuring the amount or number of some material | rāma ne **10 kilo** ālū kharīdā  ‘Ram bought ten kilo potatoes.’ |
| frequency | **vIpsA** | Measuring the frequency of occurrences of something | rāma hara 12 ghaṃṭe meṃ **do bāra** khātā hai  ‘Ram eats twice in every 12 hours.’ |

# Relative Clause relation

| **Relations** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| relative clause elaboration | **rcelab** | When the relative clause elaborates the head noun, the main verb of relative clause get this tag | hiṃda mahāsāgara jo yuropīya deśom aurā eṣiyāī deśom ko **milātā hai**, bhārata ko kendrīya sthiti pradāna karatā hai| |
| relative clause delimitation | **rcdelim** | When the relative clause delimits the head noun, the main verb of relative clause get this tag | ye aisā pha़sala hai jisakā kāma varṣā aura ucā tapamāna kī **avaśyakatā hotī ha**i| |
| relative clause cotemporal | **rcsamAnakAla** | when the temporal modifier of the subordinated event acts as the temporal modifier of the main clause event as well | jaba rāma ghara **jā rahā thā** taba bāriśa ho rahī thī |
| Relative clause colocation | **rcloc** | when the locational/spatial modifier of the subordinated event acts as the locational/ spatial modifier of the main clause event as well. | mora vahāṁ nāca rahā thā jahāṁ **bāriśa ho rahī thī** |

# Row 7: Discourse Elements

Language as a mode of communication always occurs as a discourse in which a sentence or elements within a sentence can have a connection with the previous and following sentence. This ensures cohesion and coherence in the discource. We annotate the following discourse information in this row:

* **Discourse Connective Relation-** In the next section, we discuss how we annotate intra-sentential discourse relation. See [here](#_2nuputfaxut6) for the list of discourse connective relation or sangati relation decided so far.
* **Pronominal coreference**: A discourse strategy to indicate two entities within a sentence or across sentences having the same referent.

When the antecedent of a pronoun is the whole situation and not just a noun, that pronominal expression is treated as a connective and its annotation is discussed below:

## kAryakAraNa Discourse Connective Relation

As discussed in [Sentence Simplification](#_75buigh40wu5) section, complex sentences are split into simple sentences. However, in USR we capture the connection of the split sentences in terms of some Discourse Connective Tag. This tag ensures that even after the split of a complex sentence into simple sentences, the connective information is not lost. Here is the strategy for **discourse connective annotation**:

**Case 1**: Originally the segmented sentences are connected through a single connective:

* Complex sentences are split into two simple sentences with one of them containing the connective in the sentence level

Sent\_ID: Sent\_1

* + rāma skūla nahīṃ gayā kyoṃki vaha bīmāra hai

‘Ram did not go to the school because he is sick.

* + Sent\_1a rāma skūla nahīṃ gayā ‘Ram did not go to the school’
  + Sent\_1b. kyoṃki vaha bīmāra hai **‘because** he is sick’

The USRs Sent\_1a and Sent\_1b are as follows specifying that the two sentences are connected through **kāryakārana** relation

| **Sent**\_1 | ***#***rāma skūla nahīṃ gayā kyoṃki vaha bīmāra hai | | | |
| --- | --- | --- | --- | --- |
| **Sent**\_1a | ***#***rāma skūla nahīṃ gayā | | | |
| **concept** | ramā | skūla\_1 | nahīṃ\_1 | jā\_1-yā\_1 |
| index | 1 | 2 | 3 | 4 |
|  |  |  |  |  |
| **Sent**\_1b | ***#***kyoṃki vaha bīmāra hai | | | |
| concept | wyax | bīmāra\_1 | hai\_1-pres |  |
| index | 1 | 2 | 3 |  |
| Discourse element | Sent\_1a.1: coref |  | **Sent\_1a.4:kArya kAraNa** |  |

Table 28. Single Connective in complex sentence

Notes:

* The connective present in Sent\_1b does not appear in the concept row of its USR.
* Instead, the discourse relation tag (**kArya kAraNa,** in this case) is annotated on the main verb of Sent\_1b.
* That Sent\_1b is connected to Sent\_1a is expressed by specifying the index of the main finite verb of Sent\_1a along with the relation tag.

(For the list of relations, see [Appendix 7](#_2nuputfaxut6)

Case 2: Originally the complex sentence is formed with a paired connective and one of the clauses is sub-ordinate to the other called main clause segmented sentences are connected through a paired connective

* The sentence is split into two.
* The main finite verb of the subordinate clause is chosen for discourse relation annotation
* The connective does not appear in the USR
* The discourse relation tag conveys the relation between the two segmented sentences.
* The format is: Sent\_ID.Verb\_Index:Relation\_Name

| **Sent**\_3 | ***#***yadi āpa mujhe āmaṃtrita karate haiṃ to maiṃ āpake ghara āūṃgā | | | |
| --- | --- | --- | --- | --- |
| **Sent**\_3a | ***#*** āpa mujhe āmaṃtrita karate haiṃ | | | |
| concept | addressee | speaker | āmaṃtrita+kara\_1-tā\_hai\_1 | |
| index | 1 | 2 | 3 | |
| Discourse element |  |  | **Sent**\_**3b.4:AvaSyakawA pariNAma** | |
| **Sent**\_3b | ***#*** to maiṃ āpake ghara āūṃgā | | | |
| concept | speaker | addressee | ghara\_1 | ā\_1-gā \_1 |
| index | 1 | 2 | 3 | 4 |
| Discourse element |  |  |  |  |

Table 29. Paired connective

Here is another example of complex sentences with more than one embedding:

| **Sent**\_4 | yadi āpa acchā khānā khāoge aura āpa vyāyāma karoge to āpa svastha rahoge aura āpa bīmāra nahīṃ hooge | | | |
| --- | --- | --- | --- | --- |
| **Sent**\_4a | āpa ācchā khānā khāoge | | | |
| concept | addressee | ācchā\_1 | khānā\_1 | khā\_1-gā\_1 |
| index | 1 | 2 | 3 | 4 |
| Discourse element |  |  |  |  |
| **Sent**\_4b | aura āpa vyāyāma karoge | | | |
| concept | addressee | vyāyāma+kara\_1-gā\_1 |  |  |
| index | 1 | 2 |  |  |
| Discourse element |  | **Sent**\_**4a.4:samuccaya**  **Sent**\_**4d.4:AvaSyakawA pariNAma** |  |  |
| **Sent**\_4c | to āpa svastha rahoge | | | |
| concept | addressee | svastha\_1 | raha\_1- gā\_1 |  |
| index | 1 | 2 | 3 |  |
| Discourse element |  |  |  |  |
| **Sent**\_4d | aura āpa bīmāra nahīṃ hooge | | | |
| concept | addressee | bīmāra\_1 | nahīṃ\_1 | ho\_1-gā\_1 |
| index | 1 | 2 | 3 | 4 |
| Discourse element |  |  |  | **Sent**\_**4c.3:samuccaya** |

Table 30. Paired connective with coordination and subordination

(For the list of relations, see [Appendix 7](#_2nuputfaxut6))

### Pronominal coreference

For anaphoric expression, the discourse element uses co-ref tag and also specify the index ID of the noun it corefers. The index ID of the antecedent is written as **Sent\_ID.Concept\_Index** as shown below

| **Sent**\_5 | rāma pustaka paḍha rahā hai | | | | |
| --- | --- | --- | --- | --- | --- |
| concept | rāma | pustaka\_1 | paḍha\_1-0\_ rahā\_hai\_1 | | |
| Index | 1 | 2 | 3 | | |
| Discourse element |  |  |  | | |
| **Sent**\_6 | **vaha** kala mere śahara āyā thā | | | | |
| Concept | wyax | kala\_1 | speaker | śahara | ā\_1-yā\_thā\_1 |
| Index | 1 | 2 | 3 | 4 | 5 |
| Discourse element | **sent**\_**5.1:coref** |  |  |  |  |

Table 31. Representation of Coreference in USR

# Row 8: Speaker’s view

Speaker’s view refers to the perspective of the speaker that is conveyed in the discourse through various linguistic expressions. These expressions present extra-propositional information. This row is designed in a way, where other rows, specifically concept row alone is not sufficient to capture this information or these information are not represented in concept row, such information, are captured in speaker’s view row. In this document, we are reporting some cases that we have come across in the languages which we have studied so far, namely Hindi, Bangla, English, Tamil:

1. Discourse Particles
2. Light Verbs
3. Determiners
4. Honorific, non\_honorific Pronominal forms
5. Salutation marker
6. Proximal and distal

.

What to present in Speaker’s viewpoint row-

| **Category** | **Role** | **Annotation Tag** | **Example** |
| --- | --- | --- | --- |
| Discourse Particle | Adding speaker’s evaluation of a situation or emotions- such as inclusive, exclusive, emphasis, undesirability | hI\_1, hI\_2, BI\_1 etc.  See [here](#_w7mktrkhdnq9) for more detail discussion | sūrya camakatā **bhī** hai.  ‘The Sun shines too.’ |
| Definiteness | Definiteness or specificity | def | #**beṭoṃ** ko **kheta** meṃ **bīj**a bonā cāhie. ‘The boys should sow the seed in the field.’ |
| Light Verb or raMjaka kriyā | Shade such as-volitionality, intentional, inadvertent | Shade followed by string of the verb-root and concept ID  See [Appendix 9](#_g07b95ra7vn8) for detail discussion | ramaṇa sārā miṭhāī khā **liyā**.  ‘Raman has eaten all sweets completely.’ |
| 2nd person pronominal forms | Distinction among three forms of 2nd person pronominal- tū, tūma, āpa | Respect- āpa, informal- tū, | **tū** kahāṃ rahatā hai?  ‘Where do you stay?’  **āpa** kahāṃ rahate haiṃ?  ‘Where do you stay?’ |
| Salutation marker | Respect or address | respect | pradhānamaṃtrī **jī** abhī āe haiṃ.  ‘The honorable prime minister has just arrived.’ |
| Deixis | Proximal and Distal of temporal, spatial deixis and pronominal information | proximal and distal | **yaha** kursī hai.  ‘This is a chair’ |

Table 32. Contents of Speaker’s view row

###### 

| Addressee | Informal | #तू कहाँ रहता है? #tū kahā rahatā hai? | | |
| --- | --- | --- | --- | --- |
| Concept | addressee | kim | raha\_1-tā\_hai\_1 |
| Speaker’s view row | **informal** |  |  |
| Respect | # आप कहाँ रहते हैं? # āpa kahā rahate haiṃ? | | |
| Concept | addressee | kim | raha\_1-tā\_hai\_1 |
| Speaker’s view row | **respect** |  |  |

Table 33. Representation of speaker’s view row

#### Proximal and Distal Information for wyax

The concept row represents the concept of *wyax*

* Proximal is marked for *yaha*
* Distal is marked for *vaha*

| wyax with proximal | #यह एक पुस्तक है. #yaha eka pustaka hai | | | | |
| --- | --- | --- | --- | --- | --- |
| concept | wyax | eka\_2 | pustaka\_1 | hai\_1-pres | |
| index | 1 | 2 | 3 | 4 | |
| dependency | 4:k1 | 3:quant | 4:k1s | 0:main | |
| Speaker’s view | **proximal** |  |  |  | |
| wyax with distal | vaha kitāba lāla hai | | | | |
| concept | wyax |  | kitāba | lāla\_1 | hai\_1-pres |
| index | 1 |  | 2 | 3 | 4 |
| dependency | 2:dem |  | 4:k1 | 4:k1s | 0:main |
| Speaker’s view | **distal** |  |  |  |  |

Table 34. Proximal and distal representation

# Row 9: Scope

**In development**

# Row 10: Sentence Types

Even though USRs represent the speaker’s *vivakṣā*, we have decided to maintain the type of the target sentence for ease of generation. Sentence type is declared in the 10th row.

Examples of sentence types:

| **Sentence-type** | **Example** |
| --- | --- |
| affirmative | ramā cāvala khātī hai  ‘Rama eats rice.’ |
| negative | rāma cāvala nahīṃ khātā hai  ‘Ram does not eat rice.’ |
| yn\_interrogative | kyā āpane cāvala khāyā?  ‘Did you eat rice?’ |
| interrogative | tumane cāvala kyoṃ khāe?  ‘Why did you eat rice?’ |
| imperative | āpa ghara jāo  ‘You go home.’ |
| pass\_affirmative | rāma ke dvārā rāvaṇa kā vadha kiyā gayā thā  ‘Ravana was killed by Rama.’ |
| pass\_negative | rāma ke dvārā rāvaṇa kā vadha nahīṃ kiyā gayā thā  ‘Ravana was not killed by Rama.’ |
| pass\_interrogative | rāma ke dvārā rāvaṇa kā vadha kyoṃ kiyā gayā thā?  ‘Why did Ravana was killed by Rama?’ |
| pass\_yn\_interrogative | kyārāma ke dvārā rāvaṇa kā vadha kiyā gayā thā?  ‘Was Ravana killed by Rama?’ |
| title | harī mirca kī caṭanī banāne kī vidhi:- |
| heading | #paryatana kenxra:  Followed by sentence(s)-  #Ese sWanoM para paryatana ebaM hotal vyavasAya vikasiwa ho jAwe hEM। |
| term | valita parvata:  Followed by the definition of term as an **affirmative sentence**-  hama pichale pāṭha meṃ paDa cuke haiṃ ki pṛthvī kī āntarika halacaloṃ ke kāraṇa paratadāra śailoṃ meṃ valana paDte haiṃ| |
| fragment | #nadiyoṃ tathā himakhaṃḍoṃ ke phalasvarūpa tāje jala kī āpūrti|  When the sentence occurs as a point and incomplete sentence |

Table 35. Sentence Type

# 

# Row 11: Construction

USR treats semantic frames that are linguistically expressed as *larger-than-lexical-but-smaller-than-sentential expressions* as Constructions. They are represented as the value of the Construction feature in USR.

The conceptual schema of the constructions:

(i) The Frame

(ii) The relations among the components within the frame

The characteristic features of these frames are the following:

* There is a form (more precisely a pattern) that maps to a meaning
* Each pattern is given a name that reflects the meaning
* The semantic tag of each slot of the pattern is specified in Semantic Category feature row
* The relation among the slots is also marked, when needed

Following rows in USR capture the necessary information. First, the components of the Construction are given a semantic tag in the Semantic Category row and then, the relation is specified in the Construction row. The construction is given a name **Calender**.

***#***15 agasta 1947 bhārata svādhīn huā

| Original sentence | ***#***15 agasta 1947 bhārata svādhīn huā  ‘India became independent on 15th August 1947.’ | | | | |
| --- | --- | --- | --- | --- | --- |
| Concept | 15 | agasta | 1947 | bhārata | svādhīna+ho\_1\_yā\_1 |
| Index | 1 | 2 | 3 | 4 | 5 |
| Semantic Category of Noun | dom\* | moy | yoc | place |  |
| Dependency row | 5:k7t | 1:r6 | 2:r6 | 5:k1 | 0:main |
| Construction | Calender:[1/2:in, 2/3:in] | | | | |

Table 36a. Calenderic Construction

| Original sentence | ***#***15 agasta 1947 śukrvāra rāta bhārata svādhīn huā  ‘15th August, 1947, friday night, India became independent.’ | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | 15 | agasta | 1947 | śukrvāra+rāta\_1 | bhārata | Svādhīn+ho\_1\_yā\_1 |  |
| Index | 1 | 2 | 3 | 4 | 5 | 6 |  |
| Semantic Category of Noun | dom\* | moy | yoc | dow | place |  |  |
| dependency row | 4:rs | 1:r6 | 2:r6 | 6:k7t | 6:k1 | 0:main |  |
| Construction | Calender:[1/2:in, 2/3:in] compound:[4.1/ 4.2:r6] | | | | | | |

Table 36b. Calendric Construction

Night *of* 15th *of* August *of* 1947 (‘< ‘ implies part-whole relation)

Following USR represents another construction, called-measuring construction or ***‘meas’.***

Relation between *tīna* and *kilo* is not specified by any pre or post positional marker attached with them. They together denote a quantity, which is represented in the ***semantic category of noun*** row. To show the inner compositionality between *tīna* and *kilo* we propose a measuring construction which specifies relation between two expressions as one is a **count** and the other is a **unit**.

| Sentence | ***#***rāma bājāra se tīna kilo āṭā kharīdegā. | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| Concept | rāma | bājāra\_1 | 3+kilo\_1 | | āṭā\_1 | kharīda\_1-gā\_1 |
| Index | 1 | 2 | 3 | | 4 | 5 |
| Semantic Category of Noun | per male |  | meas | |  |  |
| Morpho-Semantic Category |  |  |  | |  |  |
| Dependency Row |  | | | | | |
| Construction | meass\_meas:[3.1@count/3.2:unit:card] | | | | | |

Table 37. Measurement Construction

Conjunction or disjunction between entities or constituents is also represented in this row.

| Conjunction | ramā aura mohana skūla gaye. ‘Rama and Mohana went to school.’ | | | |
| --- | --- | --- | --- | --- |
| Concept | ramā | mohana | skūla | jā\_1-yā\_1 |
| Index | 1 | 2 | 3 | 4 |
| Construction | conj:[1,2] | | | |

Table 38. Conjunction

# Phenomena

## Spatio-directional Terms

In some languages, namely, Hindi, Bangla and other Indic languages, spatio-directional terms exhibit three roles: (a) nominal, (b) relational and (c) modificational.

A. **Nominal use**: When the spatio-directional term denotes a place.

ūpara jāo. ‘Go upstair’

All spatio-directional terms imply a directionality with reference to another place (a reference object). Thus, *upar* ‘upstairs’ specifies a locus (location) which is above a place, the reference object. When the reference object is not explicitly mentioned, it can be taken as the location of the speaker as in the above sentence. The reference object can also occur in the sentence as shown below:

peda ke ūpara cāmda hai

‘The moon is above the tree’

The moon is located in a locus that indicates a space denoted by the geometric term ūpara ‘above’ with respect to the reference object peda. The relation between peda and ūpara is discussed [here](#_9p26szwg6gqa)

B. **Relational use:**

In the following sentence, the bird is located in a locus that is in ‘part-whole’ relation with the reference object ped.a

peda ke ūpara paksı̄ hai

‘The bird is on the tree’

Here the tree is the location of the bird. ūpara is only defining the direction and therefore ke ūpara is treated as a multiword post-position. For such usage, ūpara will not appear in the concept row.

C. **Modificational use**

ghara ke sāmane bagīce meṃ eka sāṃpa hai.

‘There is a snake in the garden which is in the front side of the house’

Here, we can paraphrase the sentence as

ghara ke sāmane vāle bagīce meṃ eka sāṃpa hai.

Or

ghara ke sāmane bagīce meṃ eka sāṃpa hai.

This is a modificational use of the spatio-directional term.

Based on the multiple role played by spacio-directional terms in many languages, USR has taken following decisions to annotate them-

* When spatio-directional terms such as -ke ūpara, -ke nīca appear as a postposition in the sentence,
  + They do not appear in the concept row
  + The noun which these post-positions occur with, is given **k7p** relation
* When spatio-directional terms occur as nominal,
  + They appear in the context row
  + The referent noun they are we mark the referent object as **rdl** or relation deśalakṣaṇa and the spatio-directional term gets **k7p** relation.
* When the spatio-directional terms *upara, b*āhara*,*  pāsa etc occur in their nominal use, they are specified at concept row. See [Appendix 6](#_w5lrt3rmh8th) for all spatio-directional terms

| Postpositional use of spatio-directional term | ***#***ciḍiyā **peḍa ke** **ūpara** baiṭhī hai.  ‘The bird is sitting on the tree’ | | | |
| --- | --- | --- | --- | --- |
| Concept | ciḍiyā\_1 | **peḍa\_1** | baiṭha\_1-yā\_hai\_1 | |
| Index | 1 | 2 | 3 | |
| Dependency | 3:k1 | **3:k7p** | 0:main | |
| Nominal use of spatio-directional term | cāṁda **peḍa ke** **ūpara** hai  ‘The moon is above the tree.’ | | | |
| Concept | cāṁda\_1 | **peḍa\_1** | **ūpara\_1** | hai\_1-pres |
| Index | 1 | 2 | 3 | 4 |
| Dependency | 4:k1 | **3:rdl** | **4:k7p** | 0:main |

Table 39. Different use of Spatio-directional term

## Measuring Terms

A measuring unit is a standard quantity used to express a physical quantity. These kinds of entities specify a measurement unit, a percentage, ratio, quantity, and capacity. USR proposes following decision regarding annotation of measuring units in different layer

* Measuring unit is represented as a compound with the number in the concept row. Such as- 2+liṭara\_1
* Since measuring units are not proper nouns or do not represent any name, thus, we have considered them as a concept with ID.
* The compositionality between the components of measuring construction is specified in the Construction row.

| Measurement sentence | ***#***eka boṭala 2 liṭara vālī paipsī lānā  ‘Bring a bottle of 2-litre Pepsi.’ | | | | |
| --- | --- | --- | --- | --- | --- |
| concept | 1 | boṭala\_1 | **2+liṭara**\_1 | paipsi | lā\_1-o\_1 |
| index | 1 | 2 | 3 | 4 | 5 |
| Semantic Category |  |  | **quantity** | ne |  |
| Construction | **meas:[3.1:count,3.2:unit]** | | | | |

Table 40. Representation of Measurement

## Demonstrative

A demonstrative pronoun is a pronoun that is used for ‘pointing out the one referred to and distinguishing it from others of the same class’[Mirriam-webster]. The term ‘demonstrative’ has been used by Diessel (1999), as pronouns or noun modifiers (the ‘this/that’ kind) along with locational and temporal adverbs (the ‘here/there’, ‘now/then’ kind. Deixis is considered as one of the inherent semantic properties of demonstratives.

Demonstrative gets ‘wyax’ as a concept, and is tagged as ‘dem’ in dependency row. The proximal and distal information is encoded in the speaker's view row.

| wyax with proximal | #यह पुस्तक लाल है. #yaha pustaka lāla hai | | | | |
| --- | --- | --- | --- | --- | --- |
| concept | wyax | pustaka\_1 | lāla\_1 | hai\_1-pres | |
| index | 1 | 2 | 3 | 4 | |
| dependency | 2:dem | 4:k1 | 4:k1s | 0:main | |
| Speaker’s view | proximal |  |  |  | |
| wyax with distal | #वह किताब लाल है. # yaha kitāba lāla hai | | | | |
| concept | wyax | kitāba\_1 | | lāla\_1 | hai\_1-pres |
| index | 1 | 2 | | 3 | 4 |
| dependency | 2:dem | 4:k1 | | 4:k1s | 0:main |
| Speaker’s view | distal |  | |  |  |

Table 41. Representation of demonstrative pronoun

## Interrogative

*kim* is a root form for all interrogative words in Sanskrit. Wh-questions are treated as variables and are represented as an abstract term *kim* in USR*.* We use *kim* as a concept that represents a substitutable\_desired\_entity. See [here](#_r2chpsyqtke1) for Hindi interrogative pronoun list.

For Yes/No questions, kim does not appear in the concept row. The sentence-type row declares *yn\_interrogative* specifies the sentence-type as a yes/no interrogative sentence.

Consider a sentence-*rāma kisase ḍarate haiṃ* ‘ who is Ram afraid of’ which may have different possible answers, such as- i. Ram is afraid of some person or animal, ii. Ram is afraid of some non-animate entity, like- fire or water, and iii. Ram is afraid of some possible incident or event, like- failure in examination.

Below we show how USR encodes different information of the same interrogative sentence ‘rāma kisase ḍarate haiṃ’ considering speaker’s *vivakṣā* to represent *kim* as a variable and the variable is bound with different feature-values which helps to generate the exact form of the *kim*.

| kim | ***#***rāma kisase ḍaratā hai | | |
| --- | --- | --- | --- |
| Concept | rāma | kim | ḍara\_1-tā\_hai\_1 |
| Index | 1 | 2 | 3 |
| Answer is some animate entity- Animacy and gender and number information | per male | anim |  |
|  |  |  |
| Answer is some inanimate entity  No animacy and no gender but number information | per male |  |  |
|  |  |  |
| Answer is some event, a gerundive noun  No animacy, no GNP | per male |  |  |
|  |  |  |

Table 42. Meta rules for kim

This distinction discussed above helps us to propose a meta rule for generation of ‘kim’ words.

* When gender, number and animacy is also marked, consider the kim as a animate entity
* when gender and number are marked and animacy information is not given, it will be considered as a non-animate entity.
* When gender and number are not marked, it will be considered as an event.

## Complex Predicate

A complex predicate is a predicate which is formed by combining a noun or an adjective

with a verb.

|  | #भौतिक भूगोल भौतिक परिघटनाओं की **व्याख्या करता है** | | | | | |
| --- | --- | --- | --- | --- | --- |
| Concept | BOwika\_1 | BUgola\_1 | BOwika\_1 | pariGatanA\_1 | **vyAKyA+kara\_1-wA\_hE\_1** |
| Index | 1 | 2 | 3 | 4 | 5 |
| Dep | 2:mod | 5:k1 | 4:mod | 5:k2 | 0:main |

## 

|  | #पृथ्वी की सतह पर लगातार **परिवर्तन हो रहा है** । | | | |
| --- | --- | --- | --- | --- |
| concept | pqWvI\_1 | sawaha\_1 | lagAwAra\_1 | **parivarwana+ho\_2-0\_raha\_hE\_**2 |
| index | 1 | 2 | 3 | 4 |
| Dep rel | 2:r6 | 4:k7p | 4:krvn | 0:main |

## 

## Light verbs

In Verb-Verb complex predicates**,** light verbs carry that part of the information which ‘have a depleted semantic contribution to the event described’ (Jespersen 1965). The semantic contribution of light verbs, i.e., volitionality, intentionality etc. is captured in the Speaker's view row by adding as the shade of meaning. Thus, USR does not represent the light verb as a concept in concept row. The string of light verbs are represented in the speaker's view row, as [shade: the string of light verbs with appropriate concept ID] as shown below-

| Concept | ramaṇa | sārā | miṭhāī | khā\_1-yā\_1 |
| --- | --- | --- | --- | --- |
| Index | 1 | 2 | 3 | 4 |
| Dependency | 3:k1 | 2:mod | 4:k2 | 0:main |
| Speaker’s view |  |  |  | **[shade:le\_1]** |

The expected outcome of this given USR is- ***#***ramaṇa ne sārī miṭhāī khā lī.

Note

Verbs which cannot be used as a main verb, will not be considered as ‘light’ verbs, such as-cukA.

## Complement Clause

In USR, we consider the finite verb as head of the sentence and mark it as *0:main*. If the sentence is a complex one and there is a complement clause, then the sentence contains two finite verbs, i.e., one is the finite verb of the main clause and the other is the finite verb of the complement clause. In such cases, USR simplifies the complement clause by breaking it into two sentences.

R1-ki (Rule1-ki):

In Complement clauses, when the conjunction "ki" (that) is used, we split the sentence based on "ki." Additionally, we add the word ‘yaha’ before the verb of the sentence preceding "ki," which functions as the vākya\_karma of that sentence. Simultaneously, we remove "ki" from the simplified sentences.

Original Sentence

| Sent\_ID\_1 | #hama pichale pāṭha meṃ paDa cuke haiṃ ki pṛthvī kī āntarika halacaloṃ ke kāraṇa paratadāra śailoṃ meṃ valana paDate haiṃ।  We have studied in the last lesson how folds are formed in the rock strata by the internal earth movements. |
| --- | --- |

After sentence simplification

| Sent\_ID\_1a | #hama pichale pāṭha meṃ yaha paDa cuke haiṃ । | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| concept | speaker | | | | | | pichalA\_1 | | | | | | pāṭha\_1 | | | | | | wyax | | | | | | paDa\_1-0\_ cukA\_ hāi\_1 | | | | | |
| index | 1 | | | | | | 2 | | | | | | 3 | | | | | | 4 | | | | | | 5 | | | | | |
| Discourse element |  | | | | | |  | | | | | |  | | | | | | Sent\_ID\_1b.6:coref | | | | | |  | | | | | |
| Sent\_ID\_1b | #pṛthvī kī āntarika halacaloṃ ke kāraṇa paratadāra śailoṃ meṃ valana paDate haiṃ। | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| concept | pṛthvī\_1 | | | | | āntarika\_1 | | | | | halacala\_1 | | | | | paratadāra\_1 | | | | | śaila\_1 | | | | | valana+ paDa\_1-tā\_hāi\_1 | | | | |
| index | 1 | | | | | 2 | | | | | 3 | | | | | 4 | | | | | 5 | | | | | 6 | | | | |
| Discourse element |  | | | | |  | | | | |  | | | | |  | | | | |  | | | | |  | | | | |
| word added | **yaha** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

R2-ki (Rule2-ki):

If a sentence is segmented by the connective '*ki*' 'that', and after segmentation, the previous clause contains the object of the verb in the clause, then *yaha*  'this' is added as a modifier before the object.

In the below given example, After splitting the sentence and removing ‘कि’, the previous clause has an object ‘*jānakārī*’. As mentioned above we will add ‘*yaha*’ before the object.

Original Sentence

| Sent\_ID\_2 | #chātroṃ se bātacīta ke jarie jānakārī prāpta kījie ki kaॉleja kā vātāvaraṇa kaisā hai? |
| --- | --- |

After sentence simplification

| Sent\_ID\_2a | #chātroṃ se bātacīta ke jarie yaha jānakārī prāpta kījie। |
| --- | --- |
| Sent\_ID\_2b | #kaॉleja kā vātāvaraṇa kaisā hai? |
| word added | yaha |

## Relative Clause

Relative clauses serve the purpose of noun modification. Such as-

rāma, **jo merā bhāī hai,** saṃskṛta kā chātra hai..

‘Ram, who is my brother, is a student of Sanskrit.’

Here, the relative clause,i.e., *who is my brother* is modifying *Ram*, a noun.

In USR, relative clauses are annotated in the following ways:

* Relative pronouns are represented as *yax* in the Concept row.
* Mainly two relations between a relative clause and its head have been presently identified. They are ***delimitation*** and ***elaboration.*** The annotation tag will be ***rcdelim*** (**R**elation **C**lausal **DELIM**itation)and ***rcelab*** (**R**elation **C**lausal **ELAB**oration).
* Besides,two more specific relation tags have been identified for restrictive relative clauses for co-temporality and co-existence.They are ***rcsamAnakAla*** (**R**elation **C**lausal **samAnakAla**)for co-temporality and **rcloc** for co-existence.
* The tags **rcdelim** and **rcelab** are marked on the head of the relative clause as shown below:

rāma, **jo merā bhāī hai,** saṃskṛta kā chātra hai.

| rāma | yax | speaker | bhāī | hai | saṃskṛta | chātra | hai-pres |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|  |  |  |  | 1: rcelab |  |  |  |

The relative clause **jo merā bhāī hai** is the modifier of the noun rāma. The head of the relative clause is hai. So, the relation between the relative clause and rāma is specified under hai in the dependency row.

* The concept *yax* is co-referred to the noun it refers to.

**rcelab (Relation Clausal Elaboration):** When the relative clause adds some extra information that defines/ elaborates/ expands the modified noun, the tag ***rcelab*** is used as exemplified below.

The relative clause “**jo** yūropīya deśoṃ aura eśiyāī deśoṃ ko milātā hai” elaborates or describes hiṃda mahāsāgara

| Sent\_ID\_1 | ***#***hiṃda mahāsāgara **jo** yūropīya deśoṃ aura eśiyāī deśoṃ ko milātā hai, bhārata ko keṃdrīya sthiti pradāna karatā hai.  ‘The Indian Ocean, which joins European countries in the West and East Asian countries, gives India a central position.’ | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| concept | hiṃda+ mahāsāgara | **yax** | yūropīya deśa | eśiyāī deśa | milā\_1-tā\_hai\_1 | bhārata | keṃdrīya\_1 | sthiti\_1 | pradāna +kara\_1-tā\_hai\_1 | |
| index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| dependency | 9:k1 | 5:k1 | 5:k2 | 5:k2 | **1:rcelab** | 9:k4 | 8:mod | 9:k2 | 0:main | |
| Discourse element |  | **1:coref** |  |  |  |  |  |  |  | |

**rcdelim (Relation Clausal Delimitation):** This tag is used when the relative clause is used to identify/distinguish/restrict/spot the modified noun as is the case given below:

| Sent\_ID\_2 | ***#***ye aisī phasala hai jise kama varṣā aura ucca tāpamāna kī āvaśyakatā hotī hai.  ‘This is such a crop which requires low rainfall and high temperature. | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | wyax | aisā | phasala\_1 | hai\_1-pres | yax | kama\_1 | varṣā | ucca\_1 | tāpamāna\_1 | āvaśyakatā +ho\_1-tā\_ hai\_1 |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| dependency | 4:k1 | 3:dem | 4:k1s | 0:main | 10:k4a | 7:mod | 10:k2 | 9:mod | 10:k2 | **3:rcdelim** |
| Discourse element |  |  |  |  | 3:coref |  |  |  |  |  |

Here it is implied that crops can be of different kinds. This sentence refers to one kind of crop and the relative clause helps us to distinguish that kind. So, the relative clause delimits (or defines) the scope of the referent of the modified noun.

**rcsamAnakAla** (**R**elation **C**lausal **samAnakAla**)**:** This tag is used when the temporal modifier of the subordinated event acts as the temporal modifier of the main clause event as well.as is the case given below:

| Sent\_ID\_2 | ***#*jaba rAma Gara jA raha WA waba bariSa ho rahI WI**  ‘When Rama was going home it was raining.’ | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| Concept | yax | rAma | Gara\_1 | jA\_1-rahA\_WA\_1 | wyax | bAriSa+ho\_1-rahA\_ WA\_1 |
| Index | 1 | 2 | 3 | 4 | 5 | 6 |
| dependency | 4:k7t | 4:k1 | 4:k2p | **6:rcsamAnakAla** | 6:k7t | 0:main |
| Discourse element |  |  |  |  | 1:coref |  |

**More than one Relative Clauses is to be splitted**

When a sentence contains more than one relative clause, we split the clauses as independent clauses.

* The concept *yax* is co-referred to the noun it refers to.
* Their relation with the noun it is modifying is specified in the discourse element row, co-referring with the noun in the main clause.
* The tags **rcdelim** and **rcelab** are marked on the head of the relative clause in the discourse element row**.**

#पृथ्वी की सतह ऐसी है जिसमें पर्यावरण के तीन महत्वपूर्ण घटक आपस में मिलते हैं तथा एक दूसरे को प्रभावित करते हैं |

1. #पृथ्वी की सतह इस प्रकार की सतह है |

pqWvI 1 ne - 2:r6 - - -

sawaha\_1 2 - - 6:k1 - - -

wyax 3 - - 4:dem - - -

prakAra\_1 4 - - 5:r6 - - -

sawaha\_1 5 - - 6:k1s - - -

hE\_1-pres 6 - - 0:main - - -

2. जिसमें [पर्यावरण के तीन महत्वपूर्ण घटक आपस में मिलते हैं |

yax 1 - - 7:k7p 2:coref - -

paryAvaraNa\_1 2 - - 5:r6 - - -

wIna\_1 3 numex - 5:card - - -

mahawwapUrna\_1 4 - - 5:mod - - -

Gataka\_1 5 - - 7:k1 - - -

Apasa\_2 6 - - 7:k2 - - -

mila\_1-wA\_hE\_1 7 - - sent\_ID\_1.2:rcdelim - - -

3. waWA

जिसमें ये एक दूसरे को प्रभावित करते हैं |

yax 1 - - 4:k7p 2:coref - -

wyax 2 - - 4:k1 sent\_ID\_12.5:coref - -

eka**+**xusarA 3 - - 4:k2 - - -

praBAviwa**+**kara\_1-wA\_hE\_1 4 - - sent\_ID\_1.2:rcdelim/sent\_ID\_2.7:samuccaya - - -

| Sent\_ID\_3 | pṛthvī ke dharātala ke ūṁce uṭhe hue bhāga jinakā śikhara hajāra mīṭara se adhika ūṁcā ho aura ḍhāla tīvra ho, tathā jinake banane me lākhoṃ varṣa lage, parvata kahalāte haiṃ| | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sent\_ID\_3a | pṛthvī ke dharātala ke ūṁce uṭhe hue bhāga ko parvata kahalāte haiṃ | | | | | | |
| Concept | pṛthvī\_1 | dharātala\_1 | ūṁcā\_1 | uṭha\_1 | bhāga\_1 | parvata\_1 | kahala\_1-tā\_hāi\_1 |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Dep. Rel. | 2:r6 | 5:r6 | 4:mod | 5:rbks | 7:k2g | 7:k2 | 0:main |
| Sent\_ID\_3b | jinakā śikhara hajāra mīṭara se adhika ūṁcā ho | | | | | | |
| Concept | yax | śikhara\_1 | hajāra+mīṭara | ūṁcā\_1 | ho\_1-e\_2 |  |  |
| Index | 1 | 2 | 3 | 4 | 5 |  |  |
| Dep. Rel. | 2:r6 | 5:k1 | 4:quantmore | 5:k1s | 0:main |  |  |
| Disc. Elem. | Geo\_nios\_7ch\_0027a.6:coref |  |  |  | Geo\_nios\_7ch\_0027a.6:rcdelim |  |  |
| Sent\_ID\_3c | aura ḍhāla tīvra ho | | | | | | |
| Concept | yax | ḍhāla\_1 | tīvra\_1 | ho\_1-e\_2 |  |  |  |
| Index | 1 | 2 | 3 | 4 |  |  |  |
| Dep. Rel. | 2:r6 | 4:k1 | 4:k1s | 0:main |  |  |  |
| Disc. Elem. | Geo\_nios\_7ch\_0027a.6:coref |  |  | Geo\_nios\_7ch\_0027a.6:rcdelim |  |  |  |
| Sent\_ID\_3d | tathā jinakā banane me lākhoṃ varṣa lage | | | | | | |
| Concept | yax | bana\_1 | lākha\_1 | varṣa\_1 | laga\_1-e\_2 |  |  |
| Index | 1 | 2 | 3 | 4 | 5 |  |  |
| Dep. Rel. | 2:r6 | 5:k7 | 4:quantmore | 5:k1 | 0:main |  |  |
| Disc. Elem. | Geo\_nios\_7ch\_0027a.6:coref |  |  |  | Geo\_nios\_7ch\_0027a.6:rcdelim |  |  |

**Relative Pronoun Playing the role of Connectives**

When the relative pronoun does not refer to one specific noun but the relative clause conveys cause, purpose or result, the relative clause is split as a separate sentence. And the relative pronoun is substituted by the 3rd person pronoun ‘wyax’. Thus the following sentence will be split as shown below:

Complex sentence-

lākhoṃ varṣoṃ meṃ himālaya ke giripāda meṃ sthita bahuta baḍe besina (droṇī) meṃ jaloḍhoṃ kā nikṣepa huā, jisase isa upajāū maidāna kā nirmāṇa huā hai.

‘In the lower course, due to gentle slope, the velocity of the river decreases, which results in the formation of riverine islands.’

A. lākhoṃ varṣoṃ meṃ himālaya ke giripāda meṃ sthita bahuta baḍe besina (droṇī) meṃ jaloḍhoṃ kā nikṣepa huā

B. **isase** isa upajāū maidāna kā nirmāṇa huā hai.

| Sent\_ID\_4 | nadī ke nicale bhāgoṃ meṃ ḍhāla kama hone ke kāraṇa nadī kī gati kama ho jātī hai, jisake pariṇāmasvarūpa nadīya dvīpoṃ kā nirmāṇa hotā hai. | | | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sent\_ID\_4a | nadī ke nicale bhāgoṃ meṃ ḍhāla kama hone ke kāraṇa nadī kī gati kama ho jātī hai | | | | | | | | | | | | | | | | | | | | | | | |
| concept | nadī\_1 | | | nicalā\_1 | | | bhāga\_1 | | | ḍhāla\_1 | | | kama+ho\_1 | | | nadī\_1 | | | gati\_1 | | | kama+ho\_1-tā\_hai\_1 | | |
| index | 1 | | | 2 | | | 3 | | | 4 | | | 5 | | | 6 | | | 7 | | | 8 | | |
| dependency | 3:r6 | | | 3:mod | | | 5:k7p | | | 5:k1 | | | 8:rt | | | 7:r6 | | | 8:k1 | | | 0:main | | |
| Sent\_ID\_4b | **isake pariṇāmasvarūpa** nadīya dvīpoṃ kā nirmāṇa hotā hai. | | | | | | | | | | | | | | | | | | | | | | | |
| concept | nadī\_1 | | | | | | | | dvīpa\_1 | | | | | | | | nirmāṇa+ho\_1-tā\_hai\_1 | | | | | | | |
| index | 1 | | | | | | | | 2 | | | | | | | | 3 | | | | | | | |
| dependency | 2:r6 | | | | | | | | 3:k1 | | | | | | | | 0:main | | | | | | | |
| Discourse element |  | | | | | | | |  | | | | | | | | **Sent**\_ID\_3**a.8:kAryakAraNa** | | | | | | | |

###### 

## Discourse connective as discourse element and relation particle together

In natural language sentences it has been observed that some such connectives occur which by annotating only as a discourse connective is not sufficient enough to capture the role of its occurrence in the sentence. In such cases, we have decided to annotate them as a discourse element to capture the coherence relation whereas to annotate the relation in the speaker's view row to capture speaker’s intention or vivaksa. Consider following list of such connectives -

| Name of the connective | Discourse element information tag | Speaker’s view information tag | Example |
| --- | --- | --- | --- |
| isake atirikta | samuccaya | ke\_awirikwa | rāma āḍa़ī calātā hai| **isake atirikta,** vaha saṃgīta bhī sunatā hai| |
| isake alāvā | samuccaya | ke\_alAvA | yaha jñāna ko vistṛta karane kā prayāsa karatā hai aura ādhārabhūta saṃkalpanāoṃ ke sātha-sātha takanīkī śabdoṃ kī vyākhyā karatā hai, jo bhaugolika jñāna ke ghaṭaka haiṃ |**isake alāvā** avadhāraṇāoṃ ko kramabaddha va vyavasthita vyavahāroṃ meṃ vikasita |
| isake sāth sāth | samuccaya | samAveSI | rāma khānā banātī hai| **isake sātha-sātha**, vaha ghara kā kāma bhī karatī hai| |
| na kevala..balki | samuccaya | BI | **nā kevala** rāma paḍha़āī meṃ acchā hai **balki** khela kūda meṃ bhī bahuta āge hai| |
| itanā ki | pariNAma | iwanA\_ki | rāma khānā banātī hai| isake sātha-sātha, vaha ghara kā kāma bhī karatī hai| |

##### Sample USR

5ch\_261 #इस गीजर का फूटना इतना निश्चित समय से होता है कि लोग अपनी घडियां मिला लेते हैं।

5ch\_261 a. #इस गीजर का फूटना इतना निश्चित समय से होता है

| 5ch\_261 | #इस गीजर का फूटना इतना निश्चित समय से होता है कि लोग अपनी घडियां मिला लेते हैं। | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| 5ch\_261 a | #इस गीजर का फूटना इतना निश्चित समय से होता है | | | | | |
| concept | $wyax | gIjara\_1 | Puta\_1 | niSciwa\_1 | niSciwa\_1 | ho\_1-wA\_hE\_1 |
| index | 1 | 2 | 3 | 4 | 5 | 6 |
| Dep. rel | 2:dem | 3:r6 | 6:k1 | 5:mod | 6:k7t | 0:main |
| 5ch\_261 b | #कि लोग अपनी घडियां मिला लेते हैं | | | | | |
| concept | loga\_1 | apnA | GadZi\_1 | milA\_1-wA\_hE\_1 |  |  |
| index | 1 | 2 | 3 | 4 |  |  |
| Dep rel | 4:k1 | 3:r6 | 4:k2 | 0:main |  |  |
| Discourse element |  |  |  | 5ch\_261.7a:pariNAma |  |  |
| Speakers view |  |  |  | iwanA\_ki |  |  |

# Post position marker as discourse element and relation particle

In natural language sentences it has been observed that some such post-position markers occur which does not bring any new karaka relation but adds speaker’s vivaksa. In such cases, we have decided to give them karaka relations according to their thematic role with the mukhya visesya and capture the vivaksa expressed by those specific post-positional markers in the speaker's view row.

Some such post-position markers are- ke sAWa sAWa, ke awirikwa, ke alAvA etc.

#और आधारभूत संकल्पनाओं के साथ-साथ तकनीकी शब्दों की व्याख्या करता है।

<sent\_id= Geo\_nios\_1ch\_0003b>

#और आधारभूत संकल्पनाओं के साथ-साथ तकनीकी शब्दों की व्याख्या करता है, जो भौगोलिक ज्ञान के घटक हैं ।

$wyax 1 - - 6:k1 Geo\_nios\_1ch\_0001.2:coref proximal -

AXAraBUwa\_2 2 - - 3:mod - - -

saMkalpanA\_1 3 - pl 6:k2 - samAveSI -

wakanIkI\_1 4 - - 5:mod - - -

Sabxa\_1 5 - pl 6:k2 - - -

vyAKyA+kara\_2-wA\_hE\_1 6 - - 0:main - - -

$yax 7 - - 11:k1 3:coref/5:coref - -

BOgolika\_1 8 - - 9:mod - - -

jFAna\_4 9 - - 10:r6 - - -

Gataka\_1 10 - - 11:k1s - - -

hE\_1-pres 11 - - 5:rcelab Geo\_nios\_1ch\_0003a.4:samuccaya - -

%affirmative

\*conj:[3,5]

</sent\_id>

# In the above example, ke sAWa-sAWa could be replaced by Ora and gets karma relation with the mukya visesya of the sentence. We have annotated karma or k2 for the nominal attached with ke sAWa-sAWa and gives necessary information to capture vivaksa in the speaker's view and construction row.

# Appendix-1

## Devanagari-WX-Indic script mapping

| अ  a  a | आ  A  ā | इ  i  i | ई  I  ī | उ  u  u |
| --- | --- | --- | --- | --- |
| ऊ  U  ū | ए  e  e | ऐ  E  ai | ओ  o  o | औ  O  au |
| क्  k  k | ख्  K  kh | ग्  g  g | घ्  G  gh | ङ्  f  ng |
|
| च्  c  c | छ्  C  ch | ज्  j  j | झ्  J  jh | ञ्  **F**  ñ |
|
| ट्  t  Ṭ | ठ्  T  Th | ड्  d  Ḍ | ढ्  D  Ḍh | ण्  N  Ṇ |
|
| त्  **w**  **t** | थ्  **W**  **th** | द्  **x**  **d** | ध्  **X**  **dh** | न्  N  n |
|
| प्  p  p | फ्  P  ph | ब्  b  b | भ्  B  bh | म्  m  m |
|
| य्  y  y | र्  r  r | ल्  l  l | व्  v  v | ं  M |
|
| श्  S  ś | ष्  **R**  **Ṣ** | स्  s  s | ह्  h  h | ः  H |
|
| ऋ  q  ri | ॠ  Q  rī | ऌ  L  li |  |  |

# Appendix-2

## Concept Dictionary Entry Format

| **Sense\_Label** | **Hindi\_Label** | **Eng\_Label** | **Example** |
| --- | --- | --- | --- |
| A\_1 | A\_1 | come\_1 | राम घर आता है  rāma ghara ātā hai  “Ram comes home” |
| A\_2 | A\_2 | know\_1 | राम को हिन्दी आती है  rāma ko hindī ātī hai  “Ram knows Hindi” |
| Pala\_1 | Pala\_1 | fruit\_1 | मुझे एक फल दो  mujhe eka phala do  “Give me a fruit” |
| Pala\_2 | Pala\_5 | result\_2 | बुरे काम बुरे फल देते हैंं ।  bure kāma bure phala dete haiṃ.  “Bad deeds give bad results.” |

# Appendix - 3

Types of Noun Compound

| Types of Noun Compound | Example |
| --- | --- |
| NC with 2 or more than 2 nouns | vana\_1+ saṃrakṣaṇa\_1 |
| NC with two modifier and one head noun | skūla\_1(M)+ śikṣaka\_1(M)+samiti\_1(H) |
| NC with modifier of modifier | vanya\_1(MM)+ jīva\_1(M)+ saṃsādhana\_1(H) |
| NC with one named entity and other nominals | kalakattā(M)(ne)+ pulisa\_1(H) |

# Appendix-4

## List of Hindi interrogative pronoun

| kyā | what |
| --- | --- |
| kauna | who |
| kahāṁ | where |
| kahāṁ se | From where |
| kaise | how |
| kisase | With whom |
| kaisā | how |
| kyoṃ | why |
| kaba | when |
| kauna sā | Which one |
| kise | To whom |
| kisakā/ kisakī | whose |
| kisa liye | why |
| kitanā/kitanī | How much/ how many |

# Appendix - 5

## Hindi TAM Dictionary

| wA\_hE\_1 | 0 - | present |
| --- | --- | --- |
| 0\_rahA\_hE\_1 | be\_ing - | Present progressive |
| 0\_sakawA\_hE\_1 | can\_0 | Present modal |
| 0\_sakawA\_WA\_1 | could\_0 | Past modal |
| 0\_sakawA\_hE\_2 | might\_0 | Present modal |
| 0\_sakawA\_hE\_3 | may\_0 | pres modal |
| 0\_hE\_1 | be\_ing- | Present progressive |
| yA\_1 | ed | past |
| gA\_1 | will\_0 | future |
| gA\_2 | would\_0 | Present modal |
| 0\_rahA\_WA\_1 | was\_ing- | Past progressive |
| wA\_rahawA\_hE\_1 | keep\_ing- |  |
| yA\_jA\_yA\_WA\_1 | was\_en - | Past in passive |
| yA\_jA\_yA\_WA\_2 | had\_been\_en | Past perfect progressive |
| yA\_jA\_yA\_1 | got\_en |  |
| yA\_jA\_yA\_hE\_1 | are\_en |  |
| yA\_jA\_yA\_hE\_2 | has\_been\_en | Present perfect in passive |
| yA\_jA\_wA\_hE\_1 | is\_en |  |
| yA\_jA\_wA\_WA\_1 | was\_en |  |
| yA\_jA\_gA\_1 | will\_be\_en |  |
| yA\_hE\_1 | has\_en | Present perfective |
| wA\_WA\_1 | usedto\_0 | Habitual past |
| 0\_jAwA\_WA\_1 | usedto\_0 | Past modal |
|  |  |  |
| o\_1 |  | imperative |
| o\_2 |  | Future imperative with more polite request, such as-kIjiyegA,kariyegA |
| -e\_1 |  | Subjunctive, such as- kareM, jAyeM,KAye |
| 0\_cukA\_hogA\_1 | have\_en |  |
| 0\_cukA\_WA\_1 | had\_en | Past perfective |
| 0\_cukA\_hE\_1 | have\_en | Present perfective |
| wA\_jA\_rahA\_hE\_1 | Keep on doing X |  |
| 0\_sakA\_1 | could\_0 |  |
| 0\_rahA\_hogA\_1 |  |  |
| 0\_rahA\_hogA\_2 | shall\_be\_ing |  |
| nA\_hE\_1 | have\_to\_0 | compulsive |
| nA\_cAhie\_1 | should\_0 | Present modal/ suggestive |
| nA\_cAhie\_2 | must\_0 | Present modal |
| nA\_hogA\_1 | must\_0 | Pres modal |
| nA\_padZA\_1 | had\_to\_0 | Past modal |
| nA\_padZegA\_1 | will\_have\_to | Future modal |
| nA\_padZawA\_hE\_1 |  |  |
| nA\_padZawA\_WA\_1 |  |  |
| nA\_lagawA\_hE\_1 |  |  |
| yA\_hogA\_1 | will\_have\_en |  |
| yA\_hogA\_2 | might\_have\_en | Past modal |
| yA\_hogA\_3 | must\_have\_en | Past modal |
| -wA\_jAwA\_hE\_1 |  |  |
| -yA\_jA\_cukA\_hE\_1 |  |  |
| -AI\_xewA\_hE\_1 |  | Such as-sunAI xewA hE |
| -AI\_padawA\_hE\_1 |  | Such as-sunAI padawA hE |
| -AI\_xiyA\_1 |  | Such as-sunAI xiyA |
| **tUta\_1+Puta\_1-**wA\_hE\_1 |  | Reduplication or eco-formation of the root verb |
| kara\_1-yA\_jAe\_1 |  | **परिवर्तन किए जाएँ।** |

# Appendix- 6

## List of Spatio-directional terms in Hindi

| **Term** | **Semantics** | **Example** |
| --- | --- | --- |
| andara | Spatial inside | rāma kāra ke aṃdara baiṭhā hai.  ‘Ram is sitting inside the house.’ |
| bāhara | Spatial outside | nāriyala kā bāhara kaṭhora hotā hai.  ‘The outside of the coconut is hard.’ |
| āge | Directional ahead | kāra ke āge naṃbara pleṭa lagī huī hai.  ‘The number plate is attached infront of the car.’ |
| sāmane | Directional front facing | mere sāmane eka nayā saca khula gayā  ‘A new truth was opened infront of me.’ |
| pīche | Directional behind | mere pīche bāta mata karo.  ‘Do not talk behind me.’ |
| ūpara | Directional on | laipaṭopa ko ṭebala ke ūpara rakheṃ.  ‘Keep the laptop on the table.’ |
| nīce | Directional under | ṭebala ke nīce mūlya lebala cipakā deṃ.  ‘Stick the price tag under the table.’ |
| dāyeṃ | Directional right | kone se dāhinī ora muḍeṃ.  ‘Take the right turn from the corner.’ |
| bāyeṃ | Directional left | saḍaka ke bāīṃ ora eka kāra khaḍaī hai.  ‘A car is parked at the left side of the car.’ |
| cāroṃ ora | Directional around | bekiṃga ṭre ke cāroṃ ora thoḍaā makkhana lagāeṃ.  ‘Apply some butter around the baking tray.’ |
| bīca | Spatial between | samudra ke bīca meṃ eka nāva hai.  ‘There is a boat in the mid of the sea.’ |
| pāsa | Spatial near | unake pāsa kucha dilacaspa kahāniyāṃ haiṃ  ‘He has some interesting stories.’ |
| dūra | Spatial far | vaha dūra bhaviṣya dekha sakatā hai  ‘He can see the far future.’ |
| nikata |  |  |

# Appendix-7 List of Discourse Connectives

| **Name of discourse relation** | **Marker** | **Tag** | **Example** | **Explanation** |
| --- | --- | --- | --- | --- |
| Avaśyakatā pariṇāma | yadi…to/ agara…to/ yadi….taba | AvaSyakawApariNAma | yadi rāma āegā to maiṃ jāūṃgī.  agara rāma ātā hai to maiṃ jāūṃgī.  ‘If Ram comes then I will go.’ | The marker indicates that the occurrence or truth of one clause depends on a specific condition stated in the other clause. |
| Avaśyakatā pariṇāma.nahīm | nahīm…to/ agara…to/ to | AvaSyakawApariNAma.nahIM | rāma āegā  nahīm to maiṃ jāūṃgī.  ‘If Ram does not come then I will go.’ |  |
| virodhī | para/ lekīna/parantu/ kintu | viroXI | rāma kā ghara choṭā hai lekīna śyāma kā ghara baḍaā hai.  ‘Ram’s house is small but Shyam’s house is big.’ | Proposition or clause presents information or a viewpoint that contradicts or stands in opposition to another proposition or clause. |
| samuccaya | Ora/ evaM/ tathā | samuccaya | rāma ko seba pasaṃda hai aura mohana ko anāra pasaṃda hai.  ‘Ram loves apple and Mohana loves pomegranate | It serves to connect and coordinate elements that are grammatically equal in importance, such as words, phrases, or clauses. |
| anyawara | yā/athabā | anyawara | āpa bājāra jāeṃge yā maiṃ jāūṃ.  ‘Either you will go to the market or I will.’ | It signals that the propositions or clauses being connected are mutually exclusive or present alternative options. |
| vyabhicāra | yadyapi…tathāpi/ yadyapi…phir bhi/ isake bāvajZuda | vyaBicAra | yadyapi rāma paḍhāī meṃ acchā thā lekīna vaha pāsa nahīṃ ho sakā.  ‘Although Ram was good at studying, he could not pass.’ | It involves the expression of a concession or acknowledgment of a contrary or unexpected fact, condition, or viewpoint, while still maintaining the overall argument or main point. |
| uttarkāla | phira,  isake bAxa, bAxa meM | uwwarakAla | pahale sunūṃgā, phira likhūṃgā  ‘First I will listen, then I will write.’ | The simultaneous temporal occurrences of two events, the connective is attached with the later event. |
| kāryakāraṇa | cūṃki/kyoṃki | kAryakAraNa | rāma skūla nahīṃ gayā kyoṃki vaha bīmāra hai  ‘Ram did not go to the school because he is sick. |  |
| pariNāma | isIlie,isalie, isake pariNAmasvarUpa,isa kAraNa | pariNAma | rāma bīmāra hai  isaliye vaha skūla nahīṃ gayā  ‘Rama is sick, thus, he did not go to school.’ |  |
| samuccaya.awirikwa [where samuccaya is the discourse element tag and ke\_awirikwa is the speaker’s view tag] | isake atirikta | samuccaya | rāma āḍa़ī calātā hai| **isake atirikta,** vaha saṃgīta bhī sunatā hai| | When additional information is added to an existing one, stated before, we use samuccaya relation as discourse information and the discourse particle which brings the speaker’s view, will be represented in the speaker's view row.  Such as- for the discourse connective, isake awirikwa, samuccaya will be the relation name in discourse element row and awirikwa will be the information encoded in speaker’s view row. |
| samuccaya.alAvA [where samuccaya is the discourse element tag and ke\_alAvA is the speaker’s view tag] | isake alāvā | samuccaya | yaha jñāna ko vistṛta karane kā prayāsa karatā hai aura ādhārabhūta saṃkalpanāoṃ ke sātha-sātha takanīkī śabdoṃ kī vyākhyā karatā hai, jo bhaugolika jñāna ke ghaṭaka haiṃ |**isake alāvā** avadhāraṇāoṃ ko kramabaddha va vyavasthita vyavahāroṃ meṃ vikasita karane kā prayāsa karatā hai | |  |
| samuccaya.BI[where samuccaya is the discourse element tag and samAveSI is the speaker’s view tag] | na kevala..balki | samuccaya | **nā kevala** rāma paḍha़āī meṃ acchā hai **balki** khela kūda meṃ bhī bahuta āge hai| |  |
| samuccaya.samAveSI [where samuccaya is the discourse element tag and samAveSI is the speaker’s view tag] | isake sāth sāth | samuccaya | rāma khānā banātī hai| **isake sātha-sātha**, vaha ghara kā kāma bhī karatī hai| |  |
| virodha.dyotaka | jabaki | viroXa.xyowaka | sāmānyataḥ prākṛtika tatvoṃ jaise parvatoṃ, nadiyoṃ, jhīloṃ ādi meṃ dhīre-dhīre parivartana hotā hai **jabaki** sāṃskṛtika tatvoṃ jaise bhavanoṃ, saḍakoṃ, phasaloṃ ādi meṃ tejī se parivartana hotā hai | | When the discourse connective states a contrast between two arguments, also known as antithesis. |
| kārya.dyotaka | ताकि | kArya.xyowaka | mānacitrakāroṃ ko bhūgaṇita ke sātha-sātha ādhunika gaṇita meṃ bhī pāraṃgata honā cāhie **tāki** ve samajha sakeṃ ki pṛthvī kī ākṛti, parīkṣaṇa ke lie caurasa sataha para prakṣepita mānacitra ke cinhoṃ kī vikṛti ko kisa prakāra prabhāvita karatī hai | | When the second argument states the desired result or the expectation of the speaker of the first argument. |
| arWAwa | dūsare śabdoṃ meṃ/arWAwa | arWAwa | Geo ‘‘pṛthvī’’ aura Graphy ‘‘varṇana karanā’ bhūgola kā śābdika artha hai, jo pṛthvī ke dharātalīya satahoṃ kā varṇana karatā hai |  **dūsare śabdoṃ meṃ** bhūgola vistṛta paimāne para sabhī bhautika va mānavīya tathyoṃ kī antaḥkriyāoṃ aura ina antaḥkriyāoṃ se utpanna sthalarūpoṃ kā adhyayana karatā hai| | When the second argument shifts the content of the previous argument to a different conceptual frame or reinterpret the first argument. |
| uxAharaNasvarUpa | udāharaṇa ke lie/  udāharanasvarupa | uxAharaNasvarUpa | bhūgola kā eka anya pakṣa kṣetrīya vibhinnatā ke kāraṇoṃ ke samajhane meṃ hai ki kisa prakāra sāmājika, sāṃskṛtika, ārthika aura janāṃkikī kāraka bhautika sthala rūpa ko parivartita kara rahe hai aura mānavīya hastakṣepa ke phalasvarūpa navīna sthala rūpoṃ kā nirmāṇa ho rahā hai|  **udāharaṇa ke lie** mānava, vana yā baṃjara bhūmi kā prayoga mānavīya adhivāsa ke rūpa meṃ kara rahā hai | When the second argument provides examples, details or more information on the state of a affairs described  in the previous argument. |
| Meanwhile | isa bīca | Meanwhile | bhārata meṃ sabase pahalī ṭrena varṣa 1853 meṃ mumbī se thāṇe ke bīca calī aura 34 ki.mī. kī dūrī taya kī|  **isa bīca** bhāratīya rela taṃtra meṃ bahuta jyādā vikāsa tathā abhivṛddhi huī| | When the following argument states or adds a new fact which happens during the time of the previous argument. |
| vivaraṇa | A specific nominal. | vivaraNa | hama sabhī isa **tathya** se acchī taraha se paricita haiṃ| pṛthvī ke dharātala kā 71 pratiśata bhāga sāgara aura mahāsāgaroṃ se ghirā huā hai| | When a specific nominal of the previous argument is elaborated in the following argument, the nominal of the previous argument gets *vivaraNa* tag in discourse element row. |
| In a nutshell/ In brief/ to summarize/ In short | saṃkṣepa meṃ | InShort | hamārī saṃskṛti ne kalā ke kṣetra meṃ advitīya yogadāna diyā hai, aura yahāṃ ke maṃdira, mahala, aura citrakalā isakā sabūta haiṃ|  **saṃkṣepa meṃ**, bhāratīya saṃskṛti vividhatā aura dharmikatā kī misāla hai | | The following argument provides a shorter or brief description/ restatement of previous argument/s. |
| By the way |  | ByTheWay | rāma 12 tārīkha ko banārasa āne vālā hai  **vaise** tuma kaba ā rahe ho? | When the following argument is a turn-taker, introducing a new subject or new information in the text/ conversation, we use this tag. |

# Appendix-8

## List of Hindi Discourse Particle and evaluative dimensions

भी

1. **समुच्चय** (also): पूर्वोक्त किसी व्यक्ति, वस्तु आदि के साथ “वर्तमान” व्यक्ति, वस्तु आदि का भी संग्रह करना, in additive meaning

| **Context:** |  |
| --- | --- |
| Hin\_Geo\_ncert\_6stnd\_1ch\_0018 | sūrya, caṃdramā tathā ve sabhī vastueṁ jo rāta ke samaya āsamāna meṃ camakatī haiṃ, khagolīya piṃḍa kahalātī haiṃ. ‘The sun, the moon and all those objects shining in the night sky are called celestial bodies.’ |
| Hin\_Geo\_ncert\_6stnd\_1ch\_0019 | kucha khagolīya piṃḍa baड़e ākāra vāle tathā garma hote haiṃ. ‘Some celestial bodies are very big and hot.’ |
| Hin\_Geo\_ncert\_6stnd\_1ch\_0020 | ye gaisoṃ se bane hote haiṃ. ‘They are made up of gasses.’ |
| Hin\_Geo\_ncert\_6stnd\_1ch\_0021 | inake pāsa apanī ūṣmā tathā prakāśa hotā hai, jise ve bahuta baḍī mātrā meṃ utsarjita karate haiṃ. ‘They have their own heat and light, which they emit in large amounts.’ |
| Hin\_Geo\_ncert\_6stnd\_1ch\_0022 | ina khagolīya piṃḍoṃ ko tārā kahate haiṃ.  ‘These celestial bodies are called stars.’ |
| **Example** |  |
| Geo\_ncert\_6stnd\_1ch\_0023 | sūrya **bhī** eka tārā hai. ‘The sun is a star.’ |

1. **बलार्थे (Emphasis)**: (पूर्ववर्ती शब्द के अर्थ को बल देता है), emphasizing the meaning of previous attached word

| **Context** |  |
| --- | --- |
| Hin\_Geo\_ncert\_7stnd\_4ch\_0133 | bhārata ke pūrvī samudrī taṭa para sthita oḍiśā meṃ baṃgāla kī khāḍī se uṭhane vāle cakravātoṃ kā khatarā banā rahatā hai.  ‘Odisha, located on the eastern seacoast of India is prone to cyclones that originate in the Bay of Bengal. ‘ |
| Hin\_Geo\_ncert\_7stnd\_4ch\_0134 | 17-18 akṭūbara, 1999 ko rājya ke pāṁca ज़iloṃ meṃ cakravāta āyā.  ‘On 17-18 October 1999, cyclone hit five districts of the state. ‘ |
| Hin\_Geo\_ncert\_7stnd\_4ch\_0135 | 29 akṭūbara, 1999 ko eka anya mahācakravāta āyā, jisane rājya ke eka baḍe bhāga meṃ tabāhī macāī.’ Another supercyclone occurred on 29 October 1999, that devastated large portions of the state.’ |
| Hin\_Geo\_ncert\_7stnd\_4ch\_0136 | mukhyataḥ pavana kā vega, varṣā tathā jvārīya protkarṣa se hāniyāṁ huīṃ. ‘The damages caused were mainly due to three factors: wind velocity, rain and tidal surge.’ |
| **Example** |  |
| Geo\_ncert\_7stnd\_4ch\_0137 | 260 kilomīṭara prati ghaṃṭe taka ke vega vālī pavana 36 ghaṃṭe se **bhī** ज़​yādā samaya taka calatī rahī. ‘The winds of upto 260 km. per hour lasted for over 36 hours.’ |

| **Context** |  |
| --- | --- |
| Hin\_Geo\_ncert\_7stnd\_1ch\_0007 | kakṣā meṃ pahuṁcakara ravi ne apane śikṣaka se pūchā, paryāvaraṇa kyā hai ? ‘In the class, Ravi asked his teacher 'What is the environment?’ |
| **Example** |  |
| Geo\_ncert\_7stnd\_1ch\_0008 | jo **kucha bhī** āpa apane āsa-pāsa dekhate ho, śikṣaka ne batāyā. ‘Whatever you see in your surroundings. 'said the teacher.’ |

1. **Any :** (BI occurs with kuCa/koI in the affirmative sentence)

| Context |  |
| --- | --- |
| Hin\_Geo\_ncert\_7stnd\_4ch\_0071 | mausama, vāyumaṃḍala kī pratyeka ghaṃṭe tathā dina-pratidina kī sthiti hotī hai. ‘Weather is this hour-to-hour, day to day condition of the atmosphere. ‘ |
| Example |  |
| Geo\_ncert\_7stnd\_4ch\_0072 | ārdra evaṃ garma mausama **kisī ko bhī** ciड़ciड़ā banā sakatā hai । ‘A hot or humid weather may make one irritable. ‘ |

| Context |  |
| --- | --- |
| Hin\_Geo\_ncert\_7stnd\_4ch\_0155 | jaba jala pṛthvī evaṃ vibhinna jalāśayoṃ se vāṣpita hotā hai, to yaha jalavāṣpa bana jātā hai. ‘When water evaporates from land and different water bodies, it becomes water vapour.’ |
| Example |  |
| Geo\_ncert\_7stnd\_4ch\_0156 | vāyu meṃ **kisī bhī** samaya jalavāṣpa kī mātrā ko ‘ārdratā’ kahate haiṃ. ‘Moisture in the air at any time, is known as humidity.’ |

1. **Yet/Even then:** (BI occurs with Phira)

| **Context** |  |
| --- | --- |
| Hin\_Geo\_ncert\_11stnd\_8ch-bk1\_0029 | vāyumaṃḍala meṃ choṭe-choṭe ṭhosa kaṇoṃ ko bhī rakhane kī kṣamatā hotī hai.  ‘Atmosphere has a sufficient capacity to keep small solid particles.’ |
| Hin\_Geo\_ncert\_11stnd\_8ch-bk1\_0030 | ye choṭe kaṇa vibhinna srotoṃ jaise- samudrī namaka, mahīna miṭṭī, dhueṁ kī kālimā, rākha, parāga, dhūla tathā ulkāoṃ ke ṭūṭe hue kaṇa se nikalate haiṃ.  ‘This small particles may originate from different sources and include sea salts, fine soil, smoke-soot, ash, pollen, dust and disintegrated particles of meteors. |
| **Example** |  |
| Hin\_Geo\_ncert\_11stnd\_8ch-bk1\_0031 | dhūlakaṇa prāyaḥ vāyumaṃḍala ke nicale bhāga meṃ maujūda hote haiṃ, **phira bhī** saṃvahanīya vāyu pravāha inheṃ kāphī ūṁcāī taka le jā sakatā hai.  ‘Dust particles are generally concentrated in the lower layers of the atmosphere; yet, convection air currents may transport them to great heights.’ |

1. **Still :** (BI occurs with aBI)

| **Context** |  |
| --- | --- |
| Hin\_Geo\_ncert\_6stnd\_1ch\_0109 | brahmāṃḍa kī viśālatā kī kalpanā karanā atyadhika kaṭhina hai. ‘It is difficult to imagine how big the universe is.’ |
| **Example** |  |
| Geo\_ncert\_6stnd\_1ch\_0110 | vaijñānika **abhī bhī** isake bāre meṃ adhika se adhika jānakārī ekatra karane meṃ juṭe haiṃ.  ‘Scientists are still tryingto find out more and more about it.’ |

| Meaning | Tag |
| --- | --- |
| Samuccaya (also/Inclusive) | BI\_1 |
| Emphasis | BI\_2 |
| Any | BI\_3 |
| Yet/Even then | BI\_4 |
| Still | BI\_5 |

**ही**

1. **व्यवच्छेद (Distinction, distinguishing):** संभावित अन्य आकांक्षा का निराकरण करना, distinguishing other expectancy

| **Context** |  |
| --- | --- |
| 10stnd:Hin\_Geo\_ncert\_10stnd\_1ch\_0118 | ataḥ bhūmi eka bahuta mahattvapūrṇa prākṛtika saṃsādhana hai.  ‘Thus, land is a natural resource of utmost importance.’ |
| **Example** |  |
| 10stnd:Hin\_Geo\_ncert\_10stnd\_1ch\_0119 | prākṛtika vanaspati, vanya jīvana, mānava jīvana, ārthika kriyāeṁ, parivahana tathā saṃcāra vyavasthāeṁ bhūmi para **hī** ādhārita haiṃ.  ‘It supports natural vegetation, wild life, human life, economic activities, transport and communication systems.’ |

| **Example** |  |
| --- | --- |
| Hin\_Geo\_nios\_7ch\_0023 | माउन्ट एवरेस्ट जैसे ऊँचे एक पर्वत शिखर का निर्माण तब **ही** हो पाता है जब आन्तरिक बलों का पर्वत निर्माणकारी या जमीन को ऊपर उठाने वाला कार्य बाह्य बलों के अपरदन कार्य की अपेक्षा अधिक द्रुत गति से होता है। |

1. **दृढ़ता (fixture,fastness):** पूर्ववर्ती शब्द के अर्थ को बल देना, emphasizing the meaning of the previous word

| **Context** |  |
| --- | --- |
| Hin\_Geo\_ncert\_10stnd\_5ch\_0043 | eka khanija viśeṣa jo niścita tattvoṃ kā yoga hai, una tattvoṃ kā nirmāṇa usa samaya ke bhautika va rāsāyanika paristhitiyoṃ kā pariṇāma hai.  ‘A particular mineral that will be formed from a certain combination of elements depends upon the physical and chemical conditions under which the material forms.’ |
| **Example** |  |
| Hin\_Geo\_ncert\_10stnd\_5ch\_0044 | isake phalasvarūpa **hī** khanijoṃ meṃ vividha raṃga, kaṭhoratā, camaka, ghanatva tathā vividha krisṭala pāe jāte haiṃ.  ‘This, in turn, results in a wide range of colours, hardness, crystal forms, lustre and density that a particular mineral possesses.’ |

1. **A few : (**hI occurs with kuCa)

| **Context** |  |
| --- | --- |
| 10stnd:Hin\_Geo\_ncert\_10stnd\_5ch\_0041 | kucha caṭṭāneṃ jaise cūnā patthara - kevala eka hī khanija se banī haiṃ; lekina adhikatara caṭṭāneṃ vibhin‍na anupātoṃ ke aneka khanijoṃ kā yoga haiṃ.  ‘Some rocks, for instance limestone, consist of a single mineral only, but the majority of the rock consist of several minerals in varying proportions.’ |
| **Example** |  |
| 10stnd:Hin\_Geo\_ncert\_10stnd\_5ch\_0042 | yadyapi 2000 se adhika khanijoṃ kī pahacāna kī jā cukī hai, lekina adhikatara caṭṭānoṃ meṃ kevala **kucha hī** khanijoṃ kī bahutāyata hai.  ‘Although, over 2000 minerals have been identified, only a few are abundantly found in most of the rocks.’ |

1. **Right from** : (hI in the context of a source time)

| **Context** |  |
| --- | --- |
|  |  |
| **Example** |  |
| 11stnd-bk1:Hin\_Geo\_ncert\_11stnd\_12ch-bk1\_0129 | apane prādurbhāva **se hī** pṛthvī ne jalavāyu meṃ aneka parivartana dekhe haiṃ । |

1. **Not only** [… but also] (पूर्वोक्त या अपरोक्त वाक्य के साथ “वर्तमान” वाक्य को भी संग्रह करना ) Adding the present sentence/ utterance with the previous one

| **Context** |  |
| --- | --- |
| Hin\_Geo\_ncert\_10stnd\_3ch\_0016 | ata: jala durlabhatā atyadhika aura baḍhatī janasaṃkhyā aura usake pariṇāmasvarūpa jala kī baḍhatī māṁga aura usake asamānavitaraṇa kā pariṇāma ho sakatā hai. |
| **Example** |  |
| Hin\_Geo\_ncert\_10stnd\_3ch\_0017 | jala, adhika janasaṃkhyā ke lie gharelū upayoga meṃ **hī** nahīṃ balki adhika anāja ugāne ke lie bhī cāhie. |

| Meaning | Tag |
| --- | --- |
| व्यवच्छेद (Distinction, distinguishing) | hI\_1 |
| दृढ़ता (fixture,fastness) | hI\_2 |
| A few | hI\_3 |
| Right from | hI\_4 |
| Not only…but also | hI\_5 |
| Only | hI\_6 |

2

List of Discourse particles.

| DP NEG | DP DISCOURSE | DP INTENSIFIER | DP INJ |  |  |
| --- | --- | --- | --- | --- | --- |
| nahIM | mAno | behaxa | hAz |  |  |
| na | Pira | sabse |  |  |  |
| binA |  | bahuwa |  |  |  |
| bagera | se | kahIM |  |  |  |
| noYna | sI | awyaMwa |  |  |  |
| manA | sirPa | Ora |  |  |  |
| bajAya | kevala | awi |  |  |  |
| nA | wo | KAsA |  |  |  |
| bajAe | yAnI | sarvAXika |  |  |  |
|  | Bara | aXika |  |  |  |
|  | wakarIbana | awyaXika |  |  |  |
|  | jEse | kahIM |  |  |  |
|  | hI | iwanA |  |  |  |
|  | BI | iwanI |  |  |  |
|  | Bara |  |  |  |  |
|  | lagaBaga | kAPI |  |  |  |
|  | Axi | aXikawara |  |  |  |
|  | sA | niwAMwa |  |  |  |
|  | jI | bilakula |  |  |  |
|  | basa |  |  |  |  |
|  | karIba |  |  |  |  |
|  | iwyAxi |  |  |  |  |
|  | waka |  |  |  |  |
|  | yUz |  |  |  |  |
|  | mAwra |  |  |  |  |
|  | ekamawra |  |  |  |  |
|  | ki |  |  |  |  |
|  | jEse |  |  |  |  |
|  | sI |  |  |  |  |
|  | mAnoM |  |  |  |  |
|  | yaWA |  |  |  |  |
|  | veM |  |  |  |  |
|  | mahaja |  |  |  |  |
|  | TIka |  |  |  |  |
|  | Osawana |  |  |  |  |
|  | arWAwa |  |  |  |  |
|  |  |  | Bale hI |  |  |
|  | ekaxama |  |  |  |  |
|  | prawi |  |  |  |  |
|  | wakaribana |  |  |  |  |
|  | bI |  |  |  |  |
|  | wuraMwa |  |  |  |  |
|  | e |  |  |  |  |
|  | yA |  |  |  |  |
|  | cAhe |  |  |  |  |
|  | Bale |  |  |  |  |
|  | bA |  |  |  |  |
|  | banAma |  |  |  |  |

# Appendix 9

## List of shade or light verbs

| Light verb | Semantic role |
| --- | --- |
| jā\_1 | completion |
| jā\_2 |  |
| dāla\_1 | intensity |
| dāla\_2 |  |
| ho\_1 |  |
| pā\_1 |  |
| de\_1 |  |
| le\_1 |  |

# Appendix-10

## Construction Frame in USR

* Calendric Construction

| Original Sentence | 15 julāī 2020 somavāra rāta ko vaha paidā huā thā ‘He was born on 15th July, 2020, monday night.’ | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | 15 | julāī\_1 | 2020 | somavāra\_1 | rāta\_1 | wyax | paidā+ho\_1-yā\_ thā\_1 |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Sem. Cat. | dom | moy | yoc | dow |  |  |  |
| Dependency row | 5:r6 | 1:r6 | 2:r6 | 1:rs | 7:k7t | 7:k1 | 0:main |
| Construction | calendar:[2/3:in] calendar:[1/2:in] | | | | | | |

* Spatial Construction

| Original Sentence | bhārata meṃ meghālaya rājya ke cūne kī śailoṃ ke pradeśa meṃ ghola raṃdhra dekhane ko milate haiṃ| | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | bhārata | meghālaya | rājya\_1 | cūnā\_1 | śaila\_1 | pradeśa\_1 | ghola\_1  +raṃdhra\_1 | dekha\_1 | mila\_1-tā\_ hāi\_1 |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Sem. Cat. | place | place |  |  |  |  |  |  |  |
| Dependency row | 3:r6 | 3:rs | 4:r6 | 5:r6 | 6:r6 | 9:k7p | 9:k1 | 9:rt | 0:main |
| Construction | \*spatial:[1/3:in] | | | | | | | | |

* Span Construction [with starting and ending point]

| Original sentence | #1990 se lekara 2000 taka pragati huī. ‘The progress happened during 1990 to 2000.’ | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | 1990 | | 2000 | | | pragati\_1 | | ho\_1-yā\_1 |
| Index | 1 | | 2 | | | 3 | | 4 |
| Sem. Cat. | yoc | | yoc | | |  | |  |
| Dependency row | 4:k7t | | 4:k7t | | | 4:k1 | | 0:main |
| Construction | span:[1@start, 2@end] | | | | | | | |

* Span Construction [without either starting or ending point]

| Original Sentence | bhārata aba taka kṛṣi para nirbhara hai | | | | |
| --- | --- | --- | --- | --- | --- |
| Concept | bhārata | aba\_1 | kṛṣi\_1 | nirbhara\_1 | hai\_1-pres |
| Index | 1 | 2 | 3 | 4 | 5 |
| Sem. Cat. | place |  |  |  |  |
| Dependency row | 5:k1 | 5:k7t | 5:k7 | 5:k1s | 0:main |
| Construction | span:[@start, 2@end] | | | |  |

* Conjunction Construction

| Original Sentence | rāma aura mohana skūla gaye. ‘Rama and Mohana went to school.’ | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | ramā | mohanā | | skūla\_1 | jā\_1-yā\_1 | | | |
| Index | 1 | 2 | | 3 | 4 | | | |
| Sem. Cat. | per/male | per/male | |  |  | | | |
| Dependency row | 4:k1 | 4:k1 | | 4:k2p | 0:main | | | |
| Construction | conj:[1,2] | | | | | | | |

* Disjunction Construction

| Original Sentence | rāma roṭī yā kelā khāyegā ‘Rama will eat bread or banana.’ | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | rām | roṭī\_1 | | kelā\_1 | khā\_1-gā\_1 | | | |
| Index | 1 | 2 | | 3 | 4 | | | |
| Sem. Cat. | per/male |  | |  |  | | | |
| Dependency row | 4:k1 | 4:k2 | | 4:k2 | 0:main | | | |
| Construction | disjunct:[2,3] | | | | | | | |

* Measurement Construction

‘meas ‘ semantic category is a complex category. That means it has components. The first component is a *count* and the second component is a *unit*.

* Time Measurement

| Original sentence | rāma 10 ghaṃṭe cale | | |
| --- | --- | --- | --- |
| concept | rāma | 10+ghaṃṭā\_1 | cala\_1-yā\_1 |
| index | 1 | 2 | 3 |
| Sem. Cat. | per/male | meas |  |
| Dep. Rel. | 3:k1 | 3:dur | 0:main |
| Cxn | **time\_meas**:[2.1@count/2.2@unit:card] | | |

* Distance Measurement

| Original sentence | rāma 10 kimi cale | | |
| --- | --- | --- | --- |
| concept | rāma | 10+kimi\_1 | cala\_1-yā\_1 |
| index | 1 | 2 | 3 |
| Sem. Cat. | per/male | meas |  |
| Dep. Rel. | 3:k1 | 3:extent | 0:main |
| Cxn | dist\_meas:[2.1@count/2.2@unit:card] | | |

* Mass measurement

| Original sentence | rāma ne 3 kilo āṭā kharīdā | | | |
| --- | --- | --- | --- | --- |
| concept | rāma | 3+kilo\_1 | āṭā\_1 | kharīda\_1-yā\_1 |
| index | 1 | 2 | 3 | 4 |
| Sem. Cat. | per/male | meas |  |  |
| Dep. Rel. | 4:k1 | 3:quant | 4:k2 | 0:main |
| Cxn | mass\_meas:[2.1@count/2.2@unit:card] | | | |

Length measurement

| Original sentence | rāma 6 phīṭa laṃbā hai | | | |
| --- | --- | --- | --- | --- |
| concept | rāma | 6+phīṭa\_1 | laṃbā\_1 | hai\_1-pres |
| index | 1 | 2 | 3 | 4 |
| Sem. Cat. | per/male | meas |  |  |
| Dep. Rel. | 4:k1 | 3:quant | 4:k1s | 0:main |
| Cxn | length\_meas:[2.1@count/2.2@unit:card] | | | |

Count construction

| Original sentence | 6 cammaca tela lāo | | |
| --- | --- | --- | --- |
| concept | 6+cammaca\_1 | tela\_1 | lā\_1-o\_1 |
| index | 1 | 2 | 3 |
| Sem. Cat. | meas |  |  |
| Dep. Rel. | 2:quant | 3:k1 | 0:main |
| Cxn | count\_meas:[1.1@count/1.2@unit:card] | | |

* Depth and Temperature construction

| Original sentence | 40 kilomīṭara kī gaharāī meṃ ise 1200⁰ se. honā cāhiye | | | | | |
| --- | --- | --- | --- | --- | --- |
| concept | 40+kilomīṭara\_1 | gaharāī\_1 | $wyax | 1200+digrI+@se.\_1 | ho\_1-nā\_cāhiye\_1 |
| index | 1 | 2 | 3 | 4 | 5 |
| Sem. Cat. | meas |  |  | meas |  |
| Dep. Rel. | 2:r6 | 5:k7p | 5:k1 | 5:k1s | 0:main |
| Cxn | depth\_meas:[1.1@count/1.2@unit:card] temp\_meas:[4.1@count/4.2@unit:card] | | | | |

**Rate Construction**

* Rate of measurement of distance/ speed against time

| Original sentence | rāma 80 kimī prati ghaṃṭā dauḍatā hai | | | |
| --- | --- | --- | --- | --- |
| concept | rāma | 80+kimī \_1 | 1+ghaṃṭā\_1 | dauḍa\_1-tā \_hai\_1 |
| index | 1 | 2 | 3 | 4 |
| Sem. Cat. | per/male | meas | meas |  |
| Dep. Rel. | 4:k1 | 4:vIpsa | 2:cxnpart | 0:main |
| Cxn | rate:[3/2:in\_every] dist\_meas:[2.1@count/2.2.@unit:card] time\_meas:[3.1@count/3.2@unit:card] | | | |

* Rate of count against time

| Original sentence | rāma prati do ghaṃṭe meṃ eka bāra khātā hai | | | |
| --- | --- | --- | --- | --- |
| concept | rāma | eka+ bāra\_1 | 2+ghaṃṭā\_1 | khā\_1-tā\_hai\_1 |
| index | 1 | 2 | 3 | 4 |
| Sem. Cat. | per/male |  | meas |  |
| Dep. Rel. | 4:k1 | 4:vIpsA | 2:cxnpart | 0:main |
| Cxn | rate:[3/2:in\_every] dist\_meas:[2.1@count/2.2.@unit:card] time\_meas:[3.1@count/3.2@unit:card] | | | |

* Fraction Construction

| Original sentence | pṛthvī kā tīna cauthāī bhāga jala se ghirā huā hai  . | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| concept | pṛthvī\_1 | sataha\_1 | 3/4 | bhāga\_3 | jala\_1 | ghira\_1-yā \_huā \_hai\_1 |
| index | 1 | 2 | 3 | 4 | 5 | 6 |
| Sem. Cat. | ne |  | numex |  |  |  |
| Dep. Rel. | 2:r6 | 4:r6 | 4:card | 6:k1 | 6:k3 | 0:main |
| Cxn | fraction:[3.2/3.1:in] | | | | | |

Proposed format for Compound construction is as follows-

**dependent/head:relation name**

| Original sentence | rāma basasṭapa para khaḍa़ā hai | | |
| --- | --- | --- | --- |
| concept | rāma | basa\_1+sṭapa\_1 | khaḍā+hai\_1-pres |
| index | 1 | 2 | 3 |
| Sem. cat | per/male |  |  |
| Dep. rel. | 3:k1 | 3:k7p | 0:main |
| Cxn. | compound:[2.1 /2.2:purpose] | | |

viroXi cxn

क्या आप संसाधन संपन्न **परंतु** आर्थिक रूप से पिछड़े और संसाधन विहीन परंतु आर्थिक रूप से विकसित प्रदेशों के नाम बता सकते हैं ?

Appendix-11

## Sample USR

FAQs

* Concepts with particles-how to treat some specific cases
* #प्रसाद **हाल ही** में गाँव से आया था ।

Here, hAla+hI\_1 will be the concept

* **CotA**\_1+sA\_1
* **Cases of Compounding**
* **mote wora para** will be treated as a MWE and will be written as -mote+wora+para\_1
* **cAroM ora, cAra waraPa** will be compound as cAroM+ora\_1, cAroM+waraPa\_1 when cAra does not have the literal meaning of ‘four’, but it has the meaning of ‘all’/ ‘every’.
* ‘xono’ will be the concept for ‘both’, ‘xo’ will be the concept for ‘.two’
* eka+sAWa\_1 for ‘together’ in sentences like ‘यहां स्थल जल एवं हवा एक साथ मिलते हैं।’
* hara+roja\_1 ‘everyday’ will be compound concept
* How to decode EsA/EsI-

#पृथ्वी की सतह **ऐसी** है।

This sentence will be modified as-

#पृथ्वी की सतह **isa prakAra** है।

* When Kuxa, svayaM comes as Apa Kuxa ko XeKo, or Apa svayaM ko XeKo- Kuxa/svayaM gets k2.

However, when it comes as -#आप यह स्वयं देखें | - *svayaM* does not appear in the concept row and comes in the speaker's view row.

* When eka occurs as eka\_2, it will get ‘quant’ relation.
* Axi will occur in concept and will get relation as the other concepts attached with it are getting. Such as

naxI, parvawa Axi

Here, if naxI, parvawa gets ‘re’ relation, Axi will also get ‘re’ relation.

* aBI, saBI,kaBI will **NOT** be splitted as aba+hI
* Reduplication will get one concept as a compound

Such as

* Asa+pAsa\_1
* kaBI+kaBI\_1
* Xina+prawixina\_1
* N+morpheme making adjective
* namaka\_1+yukwa\_1
* bAxala\_1+rahiwa\_1
* rUpa, such as niyamiwa rUpa will be in concept row
* KAsakara,ViSeRakara will not be in concept row
* Apasa will get a concept ID.
* wyax will not get any information on sem. Cat and morpho-semantic row.
* **uxAharaNa ke woda para** will be treated as discourse connective
* All measuring units with ‘varga’ such as varga kimI will be frozen expression as-varga+kimI\_1
* If a cardinal number comes with a suffix and makes it an ordinal number then they will be treated as two different concepts, such as- **20th jAnuAri rAma AyegA-** here **20** and **jAnuAr**i will be two different concepts.

However, if we say, jAnuAri ke 20 wAriKa rAma AyegA, then 20 wAriKa will be considered as a MWE.

paraṃtu ye/k1 aṃtarnirbharatāoṃ ke jaṭila jāla dvārā eka taṃtra meṃ **guṁthī** huī/k1s haiṃ

#स्टेट ऑफ फोरेस्ट रिपोर्ट (2015) के अनुसार वर्ष 2013 से सघन वनों के क्षेत्र में 3775 वर्ग किमी, की वृद्धि हुई है