**UNIVERSAL SEMANTIC REPRESENTATION GUIDELINE- V.4.2.1**

[**USR: A BRIEF OUTLINE 5**](#_pl5wowpbelwx)

[Motivation Behind USR 5](#_1kdq65u1mxrs)

[**ARCHITECTURE OF USR 7**](#_9zsthwrgbwcv)

[➢ Segmentation 7](#_tmann81k4sow)

[➢ Multi-layered Representation of USR 7](#_wcp8o6am1jam)

[Format of USR 7](#_v74lbrpueyxr)

[**ORIGINAL SENTENCE 8**](#_akuqed2t9z2w)

[Sub-sentence Identification 8](#_iznnf3eju7rx)

[**SENTENCE SEGMENTATION 10**](#_75buigh40wu5)

[General Principles for Segmentation 10](#_xfj97aknxrd5)

[**LEXICO-CONCEPTUAL INFORMATION 11**](#_10rkd4ulkvgl)

[● Concept 11](#_g2qc7evc6cjl)

[Representation of Concepts in USR 12](#_f4ritkgrlkw5)

[Simple Concepts 13](#_izba6rd1v1xd)

[Complex Concepts 13](#_tb0xvf1awdko)

[Multi-word Expression (MWE) 13](#_y9xn0arxwi4b)

[Pronominal 14](#_wh3y1gk0gqyi)

[➢ Pronominal Reference to an Entity 14](#_luaxky9xelr2)

[Events 15](#_l8hfzlqneocb)

[Modifiers of Entities and Events 17](#_g6t21a7yf85a)

[Phenomena 18](#_cpiy92v5onkf)

[I. Postulation of kartā/karma 18](#_zekyouroogp0)

[● Kartā in Active Mode 18](#_xwifg6n4xh10)

[● Complement Clauses: Postulating "vaha" 19](#_67u49c7l3zbw)

[● Postulating Addressee as Kartā in Imperative Sentences 20](#_gwaitqdik1n5)

[II. Percentile Representation 20](#_smk1xpqsiscq)

[III. Spatio-directional terms 21](#_9kmw6h4kq45)

[**SEMANTIC CATEGORY 21**](#_uwlmzyjb1l9m)

[● Named Entity Categories (ENAMEX): 21](#_yy2gbohmzhqj)

[● Number Expression (NUMEX) and Time Expression (TIMEX) Categories: 21](#_uumcba4zmbyf)

[● Gender Annotation Rules 21](#_vwe2gjkxqe3t)

[Representation of Semantic Category row in USR 22](#_pdv87m2npacg)

[Semantic Category Row Information 22](#_bse13j7wiy0x)

[Named Entity 22](#_x9s46dx65ph4)

[Time entity 22](#_7w5shb657qos)

[Number entity 23](#_thprcholaf3b)

[Animacy 23](#_cejqdnfm1mlr)

[Gender 23](#_yiz8k8asia1r)

[**MORPHO-SEMANTIC 23**](#_a1jgq9b64vcw)

[Representation of Morpho-Semantic Row 23](#_j0gtzabo1efs)

[Morpho-Semantic Row Information 24](#_sz8re98s3gbf)

[**SPEAKER’S VIEW 24**](#_tytt9ellp5tr)

[Key Components of Speaker’s View 25](#_5xzhiqrusd7d)

[A. Discourse Particles 25](#_o8927nibs2ft)

[B. Light Verbs (RaMjaka Kriyā) 25](#_fm8sztbpw6uz)

[C. Definiteness 25](#_liba1jlnwpyq)

[D. Pronominal Forms & Honorifics 26](#_cp6h8mto7uzd)

[E. Salutation Markers 26](#_ra158ex3a7ie)

[F. Deixis (Proximal & Distal) 26](#_ptbvcflvhhzn)

[Second person pronominal with speaker’s view information 26](#_sr5z61fdq5lp)

[➢ When not to treat as a light verb 27](#_uyz6tah1yf51)

[**DEPENDENCY RELATION 27**](#_88pfb85s2vkb)

[Notes: 27](#_xwqcv08y97p5)

[kāraka Relations 28](#_ht6sqa4ftbxx)

[Six main kāraka relations 28](#_neikz58iy04x)

[Other kāraka relations 28](#_dzjffyb6em81)

[Relations associated with kāraka 29](#_ejdf32tdh4nt)

[kāraketara Relations 30](#_5gscvs7iy587)

[samānādhikaraṇa 30](#_4uwes27ioorh)

[bhāvalakṣaṇa 30](#_aafktufr1zyn)

[kālavācī 31](#_rly1tzearuqq)

[Spatio-temporal Information 31](#_9p26szwg6gqa)

[Intra-sentential sangati 31](#_63dr01trvx59)

[Genitive or Possessive relation between two entities 32](#_38mffdmwf72k)

[Sādrisya, vibhājana and nirdhāraṇa 32](#_4nualgwcg56h)

[Other kārakatera relation 32](#_vzq7d62cskur)

[Modifier and modified relations 33](#_i59wnrsq9dun)

[Different Measurement relations 33](#_n9h2cvq6zrj0)

[Relative Clause relation 34](#_379kk4mq2ih2)

[Phenomena 35](#_2sahyu9xfmys)

[I. se and ke sātha as post-position marker (k2 and k2as relation) 35](#_5xm91r309odf)

[II. Different dependency relation with meṃ as postposition marker 35](#_vwfiy4x90q14)

[III. Post position marker as discourse element and relation particle 37](#_vr1bqmxj40j3)

[IV. Relative Clause 37](#_v5dafzp5g513)

[1. Purpose of Relative Clauses 37](#_9rryo2ywy3a3)

[2. Annotation of Relative Clauses 37](#_59psry2xcbf6)

[2.1 Representation of Relative Pronouns 37](#_7gwuwagxc7id)

[2.2 Types of Relations Between a Relative Clause and Its Head 38](#_iql0dq6bvhqp)

[3. Annotation of Relative Clauses in USR 38](#_jro0kl3fw28d)

[3.1 Relation Clausal Elaboration (rcelab) 38](#_25sik6rsd7sj)

[3.2 Relation Clausal Delimitation (rcdelim) 38](#_e7lnrxl7v66m)

[3.3 Relation Clausal samAnakAla (rcsamAnakAla) - Co-temporality 39](#_g9pa8isn2eep)

[3.4 Relation Clausal Location (rcloc) - Co-existence 40](#_2c89bzmivbho)

[3.5 Relational Clausal Proportion (rcprop) 40](#_czbkjyh1kqtx)

[4. Handling Multiple Relative/Correlative Clauses 40](#_rgnvl24qivkx)

[More than one Relative/ Correlative Clauses is to be splitted 40](#_bv0q8tmgdxmc)

[5. Handling Embedded Relative Clauses with Resumptive Pronouns 41](#_xs9ob3t35ihs)

[V. Correlative Clauses 41](#_n96xhuc571e4)

[1.1 Characteristics of Correlative Clauses 41](#_ispsjde7i9h7)

[1.2 Annotation Strategy for Correlative Clauses 41](#_xbw0h7y2xbpd)

[**DISCOURSE ELEMENTS 42**](#_f0n2v02vbgxa)

[Pronominal Coreference 43](#_8040baifo2g1)

[How to Annotate Pronominal Coreference 43](#_yf6r3vgqdp53)

[DISCOURSE CONNECTIVE 43](#_7krmiwx41x5x)

[Annotation Strategy for Discourse Connective 44](#_rdufhefecppd)

[Case 1: Single Connective in a Complex Sentence 44](#_ro80optfchj4)

[Case 2: Paired Connectives with a Main and Subordinate Clause 44](#_4rob8f3r6j4t)

[**CONSTRUCTION 46**](#_mnenl2pe7klh)

[How are complex concepts represented? 46](#_cl585gysulrp)

[Named Entity Construction 47](#_ik5qfctc3leh)

[Compound Construction 48](#_1fqb8u8wka2)

[I. Noun Compound 48](#_3y3p0nm5uuv5)

[List of Noun compound 48](#_zec3k9wsf8ia)

[When NOT to consider as Noun Compound 49](#_c0vzqqr05nhb)

[Noun Compound Construction 49](#_cuke0b255ykf)

[II. Compound Construction 50](#_e2qne94sjl42)

[Complex Predicate 51](#_rp4qv0grd09e)

[Conjunction Construction 51](#_7bu50mor561r)

[Disjunction Construction 52](#_3pfusscq91b5)

[Calendar Construction 52](#_guaakp66nbm7)

[Spatial Construction 53](#_wy1hbof5okcd)

[Temporal Construction 53](#_1y2fyjvc9d0r)

[Span Construction 54](#_fe4iraqtub7w)

[Measurement Construction 55](#_wtpbfw3bcwqy)

[Rate Construction 55](#_jnnlw2r2gyh8)

[APPENDIX 56](#_pdvgh18zk7k)

[Segmentation 56](#_2fytqw5x614n)

[I. Discourse Units and Finite Verbs: 56](#_gp3ugez1dmy1)

[II. Relative Clauses: 56](#_ov8x0i1nfxzf)

[Relative Clause Segmentation 57](#_gierfav9ryf5)

[Case A: 57](#_cqvz1d1c8pbd)

[Case B: 58](#_yb1mnofi3ppp)

[Case C: 58](#_31guhoqfesu8)

[Case D: 58](#_ubo9j3rvyc3t)

[III. Complement Clause Segmentation 59](#_vgdpziqjkrce)

[IV. Connectives 60](#_grmwp44pcu3k)

[Concepts 61](#_5r6k87sfilsj)

[I. List of Complex Concepts: 61](#_vpnu7amsej7o)

[II. Which items do not get Concept ID in the concept row 62](#_en849ycxp4md)

[III. Which items are not presented in the concept row 62](#_wtd1g2pa7x9z)

[IV. Concept representation of all Pronouns 63](#_aglh9mb28ava)

[V. List of Hindi interrogative pronoun 63](#_5jvmzz7l7q5f)

[VI. List of Spatio-directional terms in Hindi 64](#_3zb4f02glme9)

[VII. Convention of symbols used in USR 64](#_z1ck6shjhb2z)

[VIII. List of Quantifiers in Hindi 65](#_x5kkoo7w3nol)

[Speaker’s View 66](#_wpozfknt8h)

[I. List of Discourse Particle 66](#_6cxdtaouxevs)

[II. List of shade or light verbs 67](#_ama1jvn5gzya)

[Discourse Connective 67](#_w5xelydvd4io)

[I. List of Discourse Connectives 67](#_2nuputfaxut6)

[Dependency Relation 70](#_bxjic5lv27bi)

[K1 70](#_7q39s6uhmofq)

# 

# 

# 

# 

# 

# 

# **USR: A BRIEF OUTLINE**

Universal Semantic Representation (USR) is a meaning representation framework inspired by the Indian Grammatical Tradition (IGT). It views meaning as originating in the speaker’s (or author’s) mind. Through the act of speaking or writing, the speaker uses language to convey their thoughts. Thus, any discourse (or text) can be seen as a reflection of the speaker's mental representation.

This guideline is designed to assist annotators in creating USRs for written discourse. The ultimate goal is to enable the generation of multiple natural language outputs from these USRs using Natural Language Generation (NLG) systems.

### **Motivation Behind USR**

Written text encapsulates the speaker's intention to represent a situation. A situation typically involves an event, participants, and potentially related events (causal, resultant, or associative). For example:

English- Ram and Sita spoke to their younger brother at the bus-stop.

In Hindi, this basic situation (propositional information) might be expressed as:  
*rāma aura sitā ne basa aḍḍe para bhāī ke sātha bāta kī.*

* **Main Event**: Speaking
* **Participants**: Ram and Sita as *speakers*, their brother (younger) as *passive speaker*
* **Location**: Bus stop
* **Time and Aspect**: Past perfective

Building upon this, the speaker can add additional layers of information:

* The brother is younger than Ram and Sita.
* The situation is negated (e.g., *nahīṃ* ‘not’).
* Certainty is added to the negation.
* The possibility of speaking to someone else at the bus stop remains open (expressed using *to* in Hindi).

This enriched representation could be expressed in Hindi as:  
*rāma aura sitā ne basa aḍḍe para apane choṭe bhāī ke sātha to nahīṃ bāta kī.*

Similar sentences in other languages would appear as follows:

| **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

**Scope of Negation and Certainty**

USR allows precise representation of how information is scoped. For instance:

1. The speaker is certain that Ram and Sita did not talk to their younger brother.
2. The speaker is uncertain whether Ram and Sita spoke to their younger brother.

Both cases might result in similar surface sentences, but USR enables annotators to specify the intended scope of negation and certainty.

**Discourse Relations and Cohesion**

A text often contains multiple sentences linked by discourse markers that establish logical connections. For example:

* **Hindi**: *kyoṃki usa dina unakā bhāī śahara meṃ thā hī nahīṃ.*
* **Bangla**: *kāran sedina oder bhāi sahar-e chi-lo-i nā.*

Here, *kyoṃki* (‘because’) justifies one sentence with another, while anaphoric pronouns (e.g., vaha) maintain cohesiveness. Discourse particles like *hī* add the speaker's perspective or emphasis.

USR captures such intricate details, providing a machine-readable yet human-friendly representation of meaning.

# **ARCHITECTURE OF USR**

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

The Segmentation tool aims to split complex sentences into simpler ones and also add elided elements in the segmented sentences whenever necessary.

### **Multi-layered Representation of USR**

* Lexico-Conceptual Level
  + [Concept](#_g2qc7evc6cjl): Root of the meaning unit (e.g., *go\_1*, *speak\_1*).
  + [Semantic Category](#_uwlmzyjb1l9m): Classification such as Person, Organization, Place, Time.
  + [Morpho-semantics](#_a1jgq9b64vcw): Derivational forms (e.g., comparatives, causatives).
  + [Speaker’s view](#_tytt9ellp5tr): Captures definiteness, discourse particles, deixis, and salutation markers.
* Relational
  + [Dependency](#_88pfb85s2vkb): Defines how words are combined with each other with meaning and hierarchical relation
  + [Construction](#_mnenl2pe7klh): Defines relation of non-compositional larger than one lexical units.
* Discourse
  + [Connectives](#_7krmiwx41x5x): Defines coherence between sentences (e.g., *because*, *although*).
  + [Pronominal-coreference](#_8040baifo2g1): Traces referential links within discourse.

### **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 |

# **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 original sentence 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 | #bhārata meṃ vanaspatijāta aura prāṇijāta |

**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| |

**Address**: If the sentence itself is an address, such as - name of a department, organization or some other address, we will specify ‘ADDRESS’ in sentence ID after the sentence number and declare the sentence-type as ‘address’ in sentence-type row.

| Sentence ID | Sentence |
| --- | --- |
| healthdata\_hin\_target\_002**ADDRESS** | #kriściyana meḍikala kaleja, vellora ḍipārṭameṃṭa apha yūrolajī |

# 

# **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.

### **General Principles for Segmentation**

* In general, segmented segments will be a [discourse unit](#_gp3ugez1dmy1) 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.
* [Relative Clauses](#_v5dafzp5g513) with the relative pronoun referring to a noun in the sentence are not segmented.
* [Complement clauses](#_vgdpziqjkrce) will be splitted.
* The [connectives](#_grmwp44pcu3k) are the cue words to segment the units and they remain attached with the splitted segments.
* Postulation of [anaphoric/ relative pronoun](#_8040baifo2g1)
* All the coherent relations or discourse relations among the clauses, perceived from the connectives, are annotated in the [discourse row](#_f0n2v02vbgxa) accordingly.

Example

Original Sentence

| Sent\_ID\_1 | ***#***rām bīmāra hai isalie vaha skūla nahīṃ gayā  ‘Ram is sick. Therefore he did not go to school.’ |
| --- | --- |

After sentence segmentation

| Sent\_ID\_1a | ***#***rām bīmāra hai ‘Ram is sick’ |
| --- | --- |
| Sent\_ID\_1b | #isalie vaha skūla nahīṃ gayā  ‘He did not go to the school’ |

### 

### 

### 

# **LEXICO-CONCEPTUAL INFORMATION**

Lexico-Conceptual Information which is generally expressed by atomic words, multiword expressions or derived words are captured at this level. Currently, this level has information at 4 layers in USR. These layers (or rows) are (i) Concept row; (ii) Semantic Category; (iii) Morpho-semantic and (iv) speaker’s view.

#### **Concept**

Concepts are the semantic constructs. Each entry to the concept row is an unambiguous representation of a concept. Concepts can be simple or complex.

* **Format**: Root\_of\_Concept\_ConceptID  
  **Example**:
  + *bhāī\_1 (brother)*
  + *paḍha\_1 (read)*
* **Need for Concept IDs**:  
  Concept IDs resolve ambiguity in polysemous lexemes. Each unique concept in the dictionary is mapped to its equivalent across languages.

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.***

* *paḍha\_1* (‘read’)
* *paḍha\_2* (‘study’)

**Note**: No concept ID assigned for Named Entity and Pronominals. (See [here](#_en849ycxp4md) for the detailed list of concepts which do not get any concept ID assigned).

##### **Representation of Concepts in USR**

| Original Sentence | | rāma aura sītā ne dillī ke basa sṭapa ke āsapāsa apane choṭe bhāiyoṃ se hī bāta kara liyā | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Concept** | | **rāma** | **sītā** | **[ne\_1]** | **[ne\_2]** | **[conj\_1]** | dillī | [ne\_3] | **basa\_1** | **sṭapa\_1** | **[6-tat\_1]** | **āsa+pāsa\_1** | **apanā** | **choṭā\_1** | **bhāī\_1** | **bāta\_1** | **kara\_1-yā\_1** | **[cp\_1]** |

**Concepts**:

1. [**Named Entities**](#_yy2gbohmzhqj): *rāma, sītā (persons)*; *dillī (place)*
2. [**Entity and Event Modifiers**](#_g6t21a7yf85a): *choṭā (younger)*; *bhāī (brother)*; *bāta (spoke)*; all assigned concept IDs.
3. [**Pronominals**](#_wh3y1gk0gqyi): *apanā (own)*
4. [**Complex Concepts**](#_tb0xvf1awdko): *bāta kī (spoke)*; *basa sṭapa (bus stop)*.

**Representation Format**:

* [**Root and TAM**](#_l8hfzlqneocb) separated by a hyphen (*-*) for verbs.
* [**Multiword expressions**](#_y9xn0arxwi4b) (MWEs) joined with a plus (*+*).
* [**Light verbs**](#_ama1jvn5gzya) (e.g., *le*) are excluded from the concept row.
* Numbers written in word will be converted into digits (one thousand five hundred= 1500)
* [**Discourse particles**](#_6cxdtaouxevs) (e.g., *hī*) are not represented in the concept row (see [here](#_en849ycxp4md) for the list of items not represented in the concept row)
* [**Symbols**](#_z1ck6shjhb2z) used for different concepts

##### **Simple Concepts**

A **simple concept** refers to a single, fundamental idea which cannot be broken down.

Examples:

* *mohana pedala 10 ghaṃṭe calā* (‘Mohan walked for 10 hours’):
  + *Simple*: *mohana, calā*

##### **Complex Concepts**

* A **complex concept** is an amalgamation of multiple simple concepts as components of the complex concept.
* Complex concepts are denoted within bracket [].
* Each complex concept defines a construction which is discussed in detail in the [**Construction row**](#_mnenl2pe7klh)**.**

Examples:

* *mohana pedala 10 ghaṃṭe calā* (‘Mohan walked for 10 hours’):
  + *Simple*: *pedala\_1*
  + *Complex*: *[time\_meas\_1]* (*10*, *ghaṃṭā\_1*)

**Schema for Complex Concepts**:

* **Named Entity:** *[ne\_1] (mohana)*
* **Noun Compound**: *[nc\_1]* (*ūrjā vikiraṇa*)
* **Complex Predicate**: *[cp\_1]* (*snāna kiyā*)
* **Measurement Expressions**: Represented with specific tags like *[time\_meas\_1]*, *[dist\_meas\_1]*.

**Note:** See [here](#_vpnu7amsej7o) for the list of complex concepts

##### **Multi-word Expression (MWE)**

MWEs are lexical units treated as single concepts, represented by joining components with a plus (*+*).

Examples

| **MWE type** | **Example** |
| --- | --- |
| Frozen expression, opaque,idioms, echo-words etc | **Asa+pAsa\_1 ‘**around’  **xina+prawixina\_1** ‘everyday’  **cAroM+Ora\_4** ‘everywhere’  **eka+sAWa\_1** ‘together’  **kriyA+kalApa\_1** ‘activities**’**  **eka+jEsA\_1** ‘similar’  **eka+samAna\_1** ‘equal’  **digrI+selsiyAsa\_1**  **mote+wora+para\_1**  ‘generally’  **hAla+hI\_1** ‘recently’ |
|
|

### 

### 

### 

##### **Pronominal**

Examples

| [Pronominal](#_aglh9mb28ava) | Concept row | Hindi pronominals |
| --- | --- | --- |
| 1st person | speaker | maiṃ, hama |
| 2nd person | addressee | tu, tuma, āpa |
| 3rd person | wyax | yaha, vaha,yahāṁ,vahāṁ |
| [Interrogative](#_5jvmzz7l7q5f) | kim | kahāṃ, kaba, kyā |
| Relative | yax | jo, jahāṃ, jaba |
| Reciprocal | eka+xUsarA | eka dūsare |
| Reflexive | apanA | apanā |

##### **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.

##### **Events**

* Verbs can be stative or action verbs.
* Verbs include root and TAM separated by a hyphen (*-*). TAM (Tense, Aspect, Mood) markers are polysemous and represented uniquely in the TAM dictionary.
* Default form of TAM is 3rd person singular form
* A complex predicate consists of a kriyāmūla and a kriyā. They are represented as two different concepts with specific concept ID and the complex concept [cp\_1] denotes their complex construction.
* Non-finite verbs are represented in root form without TAM specified in the concept row

**Examples:**

* Finite Verbs: *kara\_1-gā\_1* (‘will do’).
* Complex Predicates: *rāma ne snāna kiyā* (‘Ram took a bath’).
* Non-finite Verbs: *rāma ne skūla jā kara khānā khāyā* (‘Ram ate after going to school’).

**Representation of TAM**

| **Verb** | ***TAM in Concept*** | **Representation Schema** |
| --- | --- | --- |
| **karegā/ karegī/ karoge**  **kiyā thā/ kiyi thī/ kiye the**  ‘did’ | *kara\_1****-gā\_1***  *kara\_1****-yā\_thā\_1*** | The default form of TAM occurs in 3rd person singular form |
| **kareg*ā***  **‘Will do’** | *kara\_1****-gā\_1*** | The TAM string is separated from the root by '-' (hyphen) |
| **kiyā thā**  ‘did’ | *kara\_1****-yā\_thā\_1*** | The multiword TAM string is written with an underscore |
| **kara rahā hai/ kara rahī haī/ kara rahī ho**  ‘is doing’ | ***kara\_1-0****\_rahā\_hai\_1* | 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. |
| **jā/ jāo/jāiye**  ‘(you) go.’ | *jā\_1****-o\_1*** | For imperative sentences, TAM will be by default, ‘o\_1’ |
| **jāiyegā**  ‘(you) will please go’ | *jā\_1****-o\_2*** | For future imperative sentences, TAM will be by default, ‘o\_2’ |
| ***jāyeM***  ‘(let’s) go’ | *jā\_1****-e\_1*** | (inclusive…Let us) |
| ***raheM***  ‘May stay’ | *raha\_1****-e\_2*** | Default subjunctive TAM (exclude speaker) |

Example of eventualities representation

| **Stative verb** | #rāma acchā hai.  ‘Ram is good.’ | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | rāma | [ne\_1] | ācchā\_1 | **hai\_1-pres** | | | |
| **Action verb** | *#*rāma khīra khā rahā hai.  ‘Ram is eating rice-pudding’ | | | | | | |
| Concept | rāma | [ne\_1] | khīra\_1 | **khā\_1-0\_rahā\_hai\_1** | | | |
| **Complex Predicate** | #rāma ne nadī me snāna kiyā.  ‘Ram took a bath in the river.’ | | | | | | |
| Concept | rāma | [ne\_1] | nadī\_1 | **snāna\_1** | | **kara\_1-yā\_1** | **[cp\_1]** |
| **Imperative** | #tuma mujhe patra bhejo/ apa mujhe patra bhejiye  ‘You send me the letter.’ | | | | | | |
|  |  |  |  |  | |  | **bheja\_1-o\_1** |
| **Imperative (future**) | #apa mujhe patra bhejiyegā  ‘You send me the letter.’ | | | | | | |
|  |  |  |  |  | |  | **bheja\_1-o\_2** |
| **Subjunctive (inclusive)** | #āie peḍa lagāeṃ  ‘Let us plant a tree!’ | | | | | | |
|  |  |  |  |  | |  | **laga\_1-e\_1** |
| **Subjunctive (exclusive)** | bhagavāna āpa para kṛpā kare  ‘May God bless you!’ | | | | | | |
|  |  |  |  |  | |  | **kara\_1-e\_2** |
| **Non-finite verb** | #rāma ne skūla jā kara khānā khāyā  ‘Ram ate going to school.’ | | | | | | |
| Concept | rāma | [ne\_1] | skūla\_1 | **jā\_1** | | khānā\_1 | khā\_1-yā\_1 |
| **Non-finite verb** | #gāyoṃ ke duhane ke liye rāma ghara gayā  ‘Ram went home to milk the cows.’ | | | | | | |
| Concept | gāya\_1 | | duha\_1 | rāma | [ne\_1] | ghara\_1 | ja\_1-yā\_1 |

##### **Modifiers of Entities and Events**

Modifiers provide additional information (adjectives, quantifiers, or intensifiers).

* **Entity Modifiers**:
  + Adjective: *purānā (old)*
  + Quantifier: *saba (all)*
  + Intensifier: *bahuta (very)*
* **Event Modifiers**:
  + Manner Adverbs: *bhāgakara (running)*.
  + Negation: *nahīṃ (not)*.

Examples of modifier of entities

| Adjective | #rāma purāne dosta ke sātha bāta kara rahā thā. | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept row | rāma | [ne\_1] | eka\_2 | **purānā\_1** | | dosta\_1 | | | bāta\_1 | kara\_1-0\_rahā\_thā\_1 | [cp\_1]a |
| Quantifier | #saba laḍake āeṃge | | | | | | | | | | |
| Concept row | **saba\_1** | | laḍakā\_1 | ā\_1-gā\_1 | |  | | | | | |
| Intensifier | #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 | #rāma roja do seba khātā hai. | | | | | | | | | | |
| Concept | rāma | [ne\_1] | | roja\_3 | | **2** | | seba\_1 | | khā\_1-tā\_hai\_1 | |
| Ordinal Number | #rāma daśaratha ke prathama putra haiṃ | | | | | | | | | | |
| Concept | rāma | [ne\_1] | | daśaratha | [ne\_2] | **prathama\_1** | | putra\_1 | | hai\_1-pres | |

Example of modifier of events

| kriyā viśeṣaṇa (manner adverb) | #rāma bhāgakara āyā | | | |
| --- | --- | --- | --- | --- |
| Concept | rāma | [ne\_1] | **bhāga\_1** | ā\_1-yā\_1 |
| Negation | #rāma nahīṃ āeṃge | | | |
| Concept | rāma | [ne\_1] | **nahīṃ\_1** | ā\_1-gā\_1 |

#### **Phenomena**

##### **I. Postulation of kartā/karma**

###### **Kartā in Active Mode**

If the kartā (subject) is missing in the original sentence in its active mode, it will be added to the concept row.

**Example:**

Original Sentence

| Original Sentence | rāma ne eka kelā khāyā aura khelane gayā|  ‘Rama ate one banana and went to play. | | | |
| --- | --- | --- | --- | --- |

After Segmentation and Postulation

| Original Sentence | rāma ne eka kelā khāyā aura khelane gayā|  ‘Rama ate one banana and went to play. | | | | |
| --- | --- | --- | --- | --- | --- |
| sent\_1a | rāma ne eka kelā khāyā. | | | | |
| Concept row | rāma | [ne\_1] | 1 | kelā\_1 | khā\_1-yā\_1 |
| sent\_1b | aura khelane gayā. | | | | |
| After postulation | aura **vaha** khelane gayā | | | | |
| Concept row | **$wyax** | | khela\_1 | jā\_1-yā\_1 |  |

Note: In Sent\_1b, the concept ID for *vaha* ($wyax) is added even though it is absent in the original sentence.

###### **Complement Clauses: Postulating "vaha"**

In cases involving complement clauses, *vaha* (referenced as $wyax) is postulated after segmentation.

**Example:**

Original sentence

| Original Sentence | #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 segmentation and postulation

| Sent\_ID\_1a | #hama pichale pāṭha meṃ **yaha** paḍha़ā cuke haiṃ । | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| concept | speaker | | | | | | pichalā\_1 | | | | | | pāṭha\_1 | | | | | | wyax | | | | | | paḍha\_1-0\_ cukā\_ hāi\_1 | | | | | | | | | | | | | | | |
| Sent\_ID\_1b | #pṛthvī kī āntarika halacaloṃ ke kāraṇa paratadāra śailoṃ meṃ valana paḍhate haiṃ। | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| concept | pṛthvī\_1 | | | | | āntarika\_1 | | | | | halacala\_1 | | | | | paratadāra\_1 | | | | | śaila\_1 | | | | | valana\_1 | | | | | paḍha\_1-tā\_hāi\_1 | | | | | [cp\_1] | | | | |

###### **Postulating Addressee as Kartā in Imperative Sentences**

**Example:**

Original sentence

| Original Sentence | ojona gaisa kā mahatva batāie|  ‘Tell about the significance of Ozone gas.’ | | | | |
| --- | --- | --- | --- | --- | --- |

After Postulation

| Original Sentence | ojona gaisa kā mahatva batāie|  ‘Tell about the significance of Ozone gas.’ | | | | |
| --- | --- | --- | --- | --- | --- |
| After postulation | **āpa** ojona gaisa kā mahatva batāie| | | | | |
| Concept | **$addressee** | ojona\_1 | gaisa\_1 | mahatva\_1 | batā\_1-o\_1 |

##### **II. Percentile Representation**

Percentile values are rephrased for conceptual clarity as follows:

| Original use | **78%** or **78 pratiśata** | |
| --- | --- | --- |
| Rephrased | **100 bhāga meṃ 78 bhāga** | |
| Concept | **100, bhāga\_1, 78, bhāga\_1** | |

##### **III. Spatio-directional terms**

[Spatio-directional terms](#_3zb4f02glme9) in Indian languages have both nominal and relational uses. Their nominal usage is represented in the concept row.

# 

# **SEMANTIC CATEGORY**

At the lexico-conceptual level, the **Semantic Category** row provides ontological information about a concept. The **Semantic Category** row specifies the classification of a concept. Currently, four primary **Named Entity** (NE) categories are being annotated. Representation of this category in **USR (Universal Semantic Representation)** includes:

### [**Named Entity Categories**](#_x9s46dx65ph4) **(ENAMEX)**:

* + **per (Person)**
  + **org (Organization)**
  + **place (Location)**
  + **ne (Other Named Entities - e.g., movies, medicine, games, etc.)**

### **Number Expression (NUMEX) and Time Expression (TIMEX) Categories:**

* + [**TIMEX**](#_7w5shb657qos)
  + [**NUMEX**](#_thprcholaf3b)
* **Additional Categories**
  + [**Animacy**](#_cejqdnfm1mlr)
  + [**Gender**](#_yiz8k8asia1r) **(Inherent Gender Only)**

### **Gender Annotation Rules**

* Only inherent gender is marked (grammatical gender is not marked in USR).
* For speaker and addressee (singular number), gender is marked based on context.
* For plural or unknown speaker/addressee, gender is not specified.
  + Example: adhyāpaka *‘professor’* → Annotated as anim, but no gender information is provided.
* 3rd person pronominal form (wyax) does not receive animacy or gender information. Instead, this information is mapped from co-reference resolution.

### **Representation of Semantic Category row in USR**

| Original sentence | *#*arjuna banārasa meṃ 10 agasta, 2021, śukravāra śāma ko 5 baje adhyāpaka ke rūpa meṃ niyukta hue | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| concept | arjuna | | | [ne\_1] | | | banārasa | | | [ne\_2] | | | 10 | | | agasta\_1 | | | 2021 | | | [calendar\_1] | | | śukravāra\_1 | | | śāma\_1 | 5 | adhyāpaka\_1 | niyukta\_1 | ho\_1-yā\_1 | [cp\_1] |
| Semantic Category of Noun | **male** | | | **per** | | |  | | | **place** | | | **dom** | | | **moy** | | | **yoc** | | |  | | | **dow** | | |  | **clocktime** | **anim** |  |  |  |

##### **Semantic Category Row Information**

|  | **Semantic Category** | **Tag** | **Example** |
| --- | --- | --- | --- |
| Named Entity | Person name, a subset of animacy | per  (will be marked on complex concept) | rāma ‘Rama’, karabi ‘Karabi’ |
| Place/ Location name | place  (will be marked on complex concept) | dillī ‘Delhi’ |
| Organization name | org  (will be marked on complex concept) | banārasa hindū yūnivarsiṭī 'Banaras Hindu University’ |
| Names of movies, medicine, cuisine, games, disease | ne  (will be marked on complex concept) | Bhagavata Gita, Mahabharata |
| Time entity | day\_of\_week | dow | Śukravāra\_1 ‘Friday’ |
| month\_of\_year | moy | agasta\_1 ‘August’ |
| year\_of\_century | yoc | 1947, 2004, sana 2004- will be written as only 2004 |
| date\_of\_month | dom | 15th |
| Century | AD | 15 saxI, - will be written as 1500 |
| BC | 15 ī.pū will be written as 1500 |
| clock\_time | clocktime | 5+baje\_1 ‘5 o’clock’ will be written as 5 with clocktime in semcat row |
| Season of a year | season | SIta\_1 ‘winter’, basanta\_1 ‘spring’ |
| Any special day | timex | Independence Day, Christmas Day |
| Number entity | count | numex | 2 ‘two’ |
| 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.’ |

# **MORPHO-SEMANTIC**

At the [Morpho-Semantic row](?tab=t.0#heading=h.xv338poinllp) of the Lexico-Conceptual level, the speaker’s *vivakṣā* 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.

### **Representation of Morpho-Semantic Row**

| Original Sentence | mohana bacce se **kama** buddhimāna hai. | | | |
| --- | --- | --- | --- | --- |
| Concept | mohana | baccā\_1 | buddhimāna\_1 | hai\_1-pres |
| Index | 1 | 2 | 3 | 4 |
| Morpho-sem |  | **pl** | **comperless** |  |

* + - *bacce,* plurality of baby information is encoded in morpho-semantic row to encode the plurality of the root concept *baccā.*
    - *kama* is the comparative degree marker for *buddhimāna*, which is not shown in concept but marked as morpho-semantic information comparless

**Morpho-Semantic Information**

##### **Morpho-Semantic Row Information**

| 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.’ |
| comparmore | Comparative degree marker | gaṃgā yamunā se **jyādā** laṃbī hai  ‘Ganga is **longer** than Yamuna’ |
| comparless | 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 |
| xviwva | Full reduplication | **ghara-ghara** patra pahuṁcā [**ghara\_1** as concept]  ‘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’. |

| Original Sentence | mohana bacce se **kama** buddhimāna hai. | | | | |
| --- | --- | --- | --- | --- | --- |
| Concept | mohana | [ne\_1] | baccā\_1 | buddhimāna\_1 | hai\_1-pres |
| Index | 1 | 2 | 3 | 4 | 5 |
| Morpho-sem |  |  | **pl** | **comparless** |  |

# 

# 

# **SPEAKER’S VIEW**

The fourth Lexico-conceptual layer is called [Speaker’s view](#_wpozfknt8h) The **Speaker’s View** annotation provides essential insights into how a speaker’s intent, focus, and emphasis are reflected linguistically. This row represents the speaker’s intent, emphasis, and perspective. It captures elements such as definiteness, honorifics, discourse particles, light verbs, and deixis.

| Original Sentence | | rāma aura sītā ne dillī kā basa sṭapa ke āsapāsa apane choṭe bhāiyoṃ se hī bāta kara liyā | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Concept** | | rāma | sītā | [ne\_1] | [ne\_2] | [conj\_1] | dillī | [ne\_3] | basa\_1 | sṭapa\_1 | **[6-tat\_1]** | **āsa+pāsa\_1** | **apanā** | **choṭā\_1** | **bhāī\_1** | **bāta\_1** | **kara\_1-yā\_1** | **[cp\_1]** |
| **Speaker’s View** | |  |  |  |  |  |  |  |  |  | **def** |  |  |  |  |  |  |  |

### 

* + - There is a def(inite) bus stop that the speaker has in mind.
    - The speaker is certain that Ram and Sita spoke to their younger brothers only (however it might be the case that other people were present over there, but they did not speak to someone else except their younger brothers).
    - The event is bāta kara liyā ‘spoke’ where *le* is the light verb. The string of light verbs are represented in the speaker's view row, as [shade: the string of light verbs with appropriate concept ID]

### **Key Components of Speaker’s View**

#### **A. Discourse Particles**

[Discourse particles](#_6cxdtaouxevs) add emphasis, exclusivity, inclusivity, affirmation or rejection (in conversation/ question-answer) or emotion to a sentence.

#### **B. Light Verbs (RaMjaka Kriyā)**

[Light verbs](#_ama1jvn5gzya) convey shades of action such as volitionality, intentionality, and intensity.

See [here](#_uyz6tah1yf51) for a list excluding light verbs.

#### **C. Definiteness**

Definiteness marks whether an entity is known or specific to the speaker.

#### **D. Pronominal Forms & Honorifics**

Second-person pronouns carry different levels of formality.

#### **E. Salutation Markers**

Markers used for addressing people with respect.

#### **F. Deixis (Proximal & Distal)**

Indicates spatial or temporal positioning.

| **Category** | **Annotation Tag** | **Example** |
| --- | --- | --- |
| Discourse Particle | hI\_1, hI\_2, BI\_1 etc.  See here for more detail discussion | sūrya camakatā **bhī** hai.  ‘The Sun shines too.’ |
| Definiteness | def | #**beṭoṃ** ko **kheta** meṃ **bīj**a bonā cāhie. ‘The boys should sow the seed in the field.’ |
| Light Verb or raMjaka kriyā | Shade followed by string of the verb-root and concept ID  See [here](#_ama1jvn5gzya) for detail discussion | ramaṇa sārā miṭhāī khā **liyā**.  ‘Raman has eaten all sweets completely.’ |
| 2nd person pronominal forms | Respect- āpa, informal- tū, | **tū** kahāṃ rahatā hai?  ‘Where do you stay?’  **āpa** kahāṃ rahate haiṃ?  ‘Where do you stay?’ |
| Salutation marker | respect | pradhānamaṃtrī **jī** abhī āe haiṃ.  ‘The honorable prime minister has just arrived.’ |
| Deixis | proximal and distal | **yaha** kursī hai.  ‘This is a chair’  **vahāṁ** kala bārish hui.  It rained there yesterday.’ |

###### **Second person pronominal with speaker’s view information**

| 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** |  |  |

**Proximal and Distal Information for $wyax**

The concept row represents the concept of *wyax*

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

#### **When not to treat as a light verb**

Certain verb combinations contribute a single meaning rather than acting as a main + light verb. For example:

* *isa bhārī patthara ko eka sthāna se dūsare sthāna taka le jānā kaṭhina hai*
  + (Carrying this heavy stone from one place to another is difficult)
  + Representation: **le+jānā\_1**

# **DEPENDENCY RELATION**

In this row, we indicate the relationship between the head and its dependent using the following format:

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

Here, "relation" refers to the role of the dependent in relation to the head. For example, if we see **2:k2**, it means that **2** is the index of the head, and **k2** (karma) represents the dependent’s relationship with the head—i.e., the dependent is the **k2** of the head.

This row captures two types of head-dependency relations:

1. **kāraka relations** – Between verbs and their dependent nouns.
2. **kāraketara (non-kāraka) relations** –
   * Between verbs and their other non-kāraka dependents.
   * Between nouns and their modifiers.

### **Notes:**

* Kāraka relation tags begin with **‘k’**.
* The head is always marked as **0:main**.
* The relationship between a **viśeṣya** (head) and its dependents is specified in this row as **index of the head : relation tag** in the dependent’s column.
* In this document, the dependent is marked in **bold**, while the head is **underlined**.
* If a concept is part of a complex concept, we do not annotate its dependency relation in this row. Instead, those relations are annotated in the **construction row**.

## **kāraka Relations**

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

### **Six main kāraka relations**

| **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 | Other than time and location | mohana **rājanīti** **para** carcā kara rahe the  ‘Mohan was 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 1. six main kāraka

**Note** - Although karta/ karma of Complex Predicate gets a genitive marker, the dependency relation is k1/k2. See [Complex Predicate](#_rp4qv0grd09e) for details.

### **Other kāraka relations**

| **Relation Name** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| *anubhava-kartā* | k4a | Experiencer | **rāma ko bukh**āra hai  ‘**Ram** has fever.’ |
| *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.’ |
| *kartā* *samanadhikarana* | k1s | kartā and its viśeṣaṇa resides in the same locus, when the verb is copulative | rāma **buddhimāna** hai  ‘Ram is intelligent.’ |
| *g*a*uṇa karma* | k2g | Secondary object | rāma ne mohana ko **ek bāte** kahī.  ‘Ram has told Mohana something.’ |
| *destination* | k2p | Destination or goal | rāma **ghara** gayā  ‘Ram went home.’ |
| *karma samanadhikarana* | k2s | karma and its viśeṣaṇa resides in the same locus | rāma mohana ko **buddhimāna** samajhatā hai  ‘Ram considers Mohan to be intelligent.’ |
| *prakṛti apādāna* | k5prk | Source material | jūte **camaḍe** **se** banate haiṃ  ‘Shoes are made of leather.’ |
| *accordingly* | k7a | according to someone/ something | **rāma ke anusāra** sītā ghara para nahīṃ hai  ‘According to Ram Sita is not at home.’ |
| *kāraka associates* | k\*as | This relation is used when one entity is related or associated with the other entity and either both entities participate directly in the event/ state of being or both fulfill the valency of the verb. ‘**ke sātha/ se saṃbaṃdhita**’ comes as a marker here to indicate the relation. | phūla ko **ḍālī ke sātha** phevikāla se joḍo  ‘Fix the flower with the branch with fevicol.’ |

Table 2. 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* | rāma  **cammaca ke sātha** kāṁṭe se sabjī khā rahā hai  ‘Ram is eating vegetables with a fork along with a spoon.’ |
| rask4 | *associate of saṃpradāna* | rāma **guru jī ke sātha** śiṣyoṃ ko dakṣiṇā detā hai  ‘Along with the honorable Guru, Ram 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.’ |
| *Relation virodhī kāraka* | rviroXIk\* | This relation is used when there is a comparison between property of two entities with **ke viparīta/ ke viruddha/ ke vipakṣa** a**s** a post-positional marker to indicate them | maidāna **pahāḍa ke viparīta** samatala bhūmi hai  ‘The plain is flat ground opposite to the hill.’ |
| *Negation in Associatives* | rasnegk\* | When there is absence of some participant | **jala ke binā** koī bhī jīva jīvita nahīṃ raha sakatā|  ‘No animal can be alive without water.’ |

Table 3.Relations associated with kāraka

## **kāraketara Relations**

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 4. 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 | rblak | VN precedes VM | sūrya **ugane ke bāda** khānā khāo  ‘Eat after the sun rises.’ |
| bhāvalakṣaṇa anantarakālika | rblpk | VN follows VM | sūrya **ugane se pahale** nahāo  ‘Bathe before the sun sets.’ |

Table 5. 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 6. 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 7. 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 |  #dakṣiṇī dolana va ela nino kī saṃyukta ghaṭanā ko īeneso ke **nāma se** jānā jātā hai | |

Table 8. 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 9. 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 10. 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.’ |
| 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.’ |
| Relation Proportion | rprop | When two entity or events are in a relation with each- other in a proportion | isa maṇḍala kī ūṁcāī **baḍhane** ke sātha-sātha tāpamāna meṃ kamī hotī jātī hai|  ‘As the height of this zone increases, the temperature decreases.’ |
| 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 11. 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](#_x5kkoo7w3nol) | 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 12. Modifier and modified relations

### **Different Measurement relations**

| **Relations** | **Tag** | **Definition** | **Example** |
| --- | --- | --- | --- |
| Measurement | **rmeas** | Measurement of verb | rāma **10 kimi** calā  ‘Ram walked ten km.’  rāma **10 ghaṃṭe** calā  ‘Ram walked for ten hours.’  tāpamāna **10 ḍigrī** ho gyā hai  ‘Temperature decreased by 10 degrees.’  rāma ne **10 kilo** ālū kharīdā  ‘Ram bought ten kilo potatoes.’ |
| duration | **dur** | Duration of some eventualities | apaharaṇa meṃ śāmila 7 aparādhī **24 ghaṃṭe ke aṃtargata** giraphtāra huā|  ‘7 convicts associated with the crime were arrested withing 24 hrs.’ |

Table 13. Different Measurement Relations

### **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** |

Table 14. Relative Clause Relation

### **Phenomena**

##### **I. se and ke sātha as post-position marker (k2 and k2as relation)**

**k2**

When the verb of *dviṣṭhaphalavyāpārātmikā* (dual resultant activities)is used with either **'se'** (with) or '**ke sātha**' (along with), we assign the k2 relation. Such actions which require two people (or conscious beings) to participate to perform the actions and the result is also equally shared by both of the participants are called *dviṣṭhaphalavyāpārātmikā*. Examples of such actions include *marry, fight, argue, converse, debate, trade, mingle* etc.

| rāma **hari se/ ke sātha** laḍāī karatā hai|  ‘Ram fights with Hari.’ |
| --- |

In the above example, the action of laḍāī karanā requires two participants to fight, i.e.- one cannot fight alone, for fighting two participants (atleast) are required. Thus, the doer *rāma* is assigned k1, whereas the other participant, *hari* is assigned k2 relation.

**k2as**

When the result of an action is seen in an object, it is considered a regular object (sāmānya karma). When a sentence contains an action whose result is visible in another object besides the main object, and that object is desired (ipsita) but not most desired (ipsitatama) by the subject, it is still considered an object. However, to convey an associative meaning from the speaker’s perspective, the postposition **ke sātha** (along with) is used. Therefore, for such cases in Universal Semantic Representation (USR), the relation "k2as" is assigned.

| tuma phūla ko **ḍālī ke sātha** joḍo  ‘Attach the flower with the bouquet.’ |
| --- |

Here, the result of the verb *joḍanā* ‘attach’ is visible on the flower along with associate object *ḍālī* ‘bouquet’. Thus, *ḍālī* is assigned k2as relation here.

##### **II. Different dependency relation with meṃ as postposition marker**

* Wherever the postposition *meṃ* (in) is used and it is forming a context (subject), the relation **'k7**' should be applied there.

| āja skūla **jāne meṃ** dera ho gayī  ‘I got late going to school today.’ |
| --- |

* Wherever the postposition *meṃ* (in) is used with the reciprocal pronoun (āpasa, eka dūsare), the relation '**k2**' should be applied there.

| kinhīṃ niścita diśāoṃ meṃ eka prakāra kī śaila cakrīya tarīke se **eka-dūsare meṃ** parivartita ho jāte haiṃ  ‘Rocks of a certain type cyclically transform into one another in specific directions.’ |
| --- |

* Wherever the postposition *meṃ* is used and it is inserted with the nominal part of the complex predicate, it would be considered as  **Complex Predicate**.

| ravi ko yaha **samajha meṃ āyā**  ‘Ravi understood this.’ |
| --- |

* Wherever the postposition *meṃ* is used and it is indicating the manner of the action, the relation '**krvn**' should be applied there.

| āpa nimnalikhita praśnoṃ ke uttara **saṃkṣepa meṃ** dījie  ‘Please provide brief answers to the following questions.’ |
| --- |

* Wherever the postposition *meṃ* is used and it is indicating the modality of the sentence, the relation '**vkvn**' should be applied there.

| **vāstava meṃ** āga kī taraha lāla dravita maigmā hī lāvā hai  ‘Indeed, lava is the molten magma that is as red as fire.’ |
| --- |

##### **III. 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 sātha sātha**, **ke awirikwa, ke alAvA** etc.

| bhūgola ādhārabhūta **saṃkalpanāoṃ ke sātha-sātha** takanīkī śabdoṃ kī vyākhyā karatā hai|  ‘Geography explains basic concepts along with technical terms.’ |
| --- |

##### **IV. Relative Clause**

###### **1. Purpose of Relative Clauses**

Relative clauses serve the purpose of modifying nouns. For example:

**Example:**

| *rāma,* ***jo merā bhāī hai****, saṃskṛta kā chātra hai.*  *Ram, who is my brother, is a student of Sanskrit.* |
| --- |

In this sentence, the relative clause *jo merā bhāī hai* modifies the noun *rāma*.

###### **2. Annotation of Relative Clauses**

###### **2.1 Representation of Relative Pronouns**

* Relative pronouns are represented as **yax** in the **Concept row**.

###### **2.2 Types of Relations Between a Relative Clause and Its Head**

Two primary relations are used to annotate relative clauses:

1. **rcdelim (**Relation Clausal **DELIMitation)**
2. **rcelab (**Relation Clausal **ELABoration)**

Additionally, three more specific tags are used for restrictive relative clauses:

* **rcsamAnakAla** (Relation Clausal **samAnakAla**) – for co-temporality.
* **rcloc** (Relation Clausal **location**) - for co-existence
* **rcprop (**Relation Clausal **proportion)**

###### **3. Annotation of Relative Clauses in USR**

###### **3.1 Relation Clausal Elaboration (rcelab)**

This tag is used when the relative clause adds extra information that defines, elaborates, or expands the modified noun.

**Example:**

| Sent\_ID\_1 | rāma, **jo merā bhāī hai,** saṃskṛta kā chātra hai | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| concept | rāma | [ne\_1] | yax | speaker | bhāī | hai | saṃskṛta | chātra | hai\_1-pres |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Dependency |  | 9:k1 | 6:k1 | 5:rhh | 6:k1s | **2: rcelab** | 8:r6 | 9:k1s | 0:main |

**Annotation:**

* The relative clause ***jo merā bhāī hai*** elaborates or describes rāma
* The relation **rcelab** is marked on the head of the relative clause.

###### **3.2 Relation Clausal Delimitation (rcdelim)**

This tag is used when the relative clause is necessary to identify or distinguish the modified noun.

**Example:**

| Sent\_ID\_2 | ***#***ye aisī phasala hai jise kama varṣā 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 | āvaśyakatā \_1 | ho\_1-tā\_ hai\_1 |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| dependency | 4:k1 | 3:dem | 4:k1s | 0:main | 9:k4a | 7:mod | 8:r6 | 9:k1 | **3:rcdelim** |
| Discourse element |  |  |  |  | 3:coref |  |  |  |  |

**Annotation:**

* The relative clause ***jise kama varṣā kī āvaśyakatā hotī hai*** restricts the noun *phasala*.
* The tag **rcdelim** is marked on the head of the relative clause.

###### **3.3 Relation Clausal samAnakAla (rcsamAnakAla) - Co-temporality**

This tag is used when the temporal modifier of the subordinated event acts as the temporal modifier of the main clause event.

**Example:**

| Sent\_ID\_2 | #tanāva taba paidā hote haiṃ jaba śaktiyāṃ do viparīta diśāoṃ meṃ dharātala ke samānāṃtara kārya karatī haiṃ| | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | wanAva\_1 | $wyax | pExA\_1 | ho\_1-wA\_hE\_1 | [cp\_1] | $yax | Sakwi\_1 | 2 | viparIwa\_1 | xiSA\_1 | XarAwala\_1 | samAnAMwara\_1 | kArya\_1 | kara\_1-wA\_hE\_1 | [cp\_2] |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Dep. Rel | 5:k1 | 5:k7t |  |  | 0:main | 15:k7t | 15:k1 | 9:card | 10:mod | 15:rd | 12:r6 | 15:krvn |  |  | **2:rcsamAnaakAla** |
| Discourse element |  |  |  |  |  | 2:coref |  |  |  |  |  |  |  |  |  |

**Annotation:**

* The relative clause ***jaba śaktiyāṃ do viparīta diśāoṃ meṃ dharātala ke samānāṃtara kārya karatī haiṃ*** provides a co-temporal context.
* The tag **rcsamAnakAla** is marked accordingly.

###### **3.4 Relation Clausal Location (rcloc) - Co-existence**

This tag is used when the location modifier of the subordinated event acts as the location modifier of the main clause event.

**Example:**

| Sent\_ID\_3 | #āpa una kṣetroṃ ke nāma batāiye jahāṁ nikṣepaṇa hotā hai| | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| concept | $addressee | $wyax | kRewra\_1 | nAma\_1 | bAwA\_1-o\_1 | $yax | nikRepaNa+ho\_1-wA\_hE\_1 |
| index | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Dep rel | 5:k1 | 3:dem | 4:r6 | 5:k2 | 0:main | 7:k7p | 3:rcloc |
| Discourse element |  |  |  |  |  | 3:coref |  |

**Annotation:**

* The relative clause ***jahāṁ nikṣepaṇa hotā hai*** specifies a location.
* The tag **rcloc** is marked accordingly.

###### **3.5 Relational Clausal Proportion (rcprop)**

This tag is used when two entities or events in two clauses, i.e.- main clause and relative clause are in a proportionate relation.

Example:

###### 

###### **4. Handling Multiple Relative/Correlative Clauses**

###### [**More than one Relative/ Correlative Clauses is to be splitted**](https://docs.google.com/document/u/1/d/1zwuDr2ZzScPrH2NUFm_USuL3bwT-MXCvZ3f1MkxU9-s/edit)

* When a sentence contains more than one relative clause, they are split as independent clauses.
* The concept **yax** is co-referred to the noun it modifies.
* Their relation with the noun is specified in the **Discourse Element Row** as a **coreference** with the noun in the main clause.
* The tags **rcdelim** and **rcelab** are marked on the head of the relative clause.

###### **5. Handling Embedded Relative Clauses with Resumptive Pronouns**

* In some cases, a resumptive pronoun is repeated in the main clause.
* In such cases, the concept of the resumptive pronoun is **not** represented in USR.

Example:

| Sent\_ID\_4 | parvata, jo jvālāmukhī se nikale padārthoṃ se bane haiṃ unheṃ saṃgrahita parvata kahate haiṃ| | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | paravawa\_1 | $yax | jvAlAmuKI\_1 | nikala\_1 | paXArWa\_1 | bana\_1 | hE\_1-pres | saMgrahiwa\_1 | paravawa\_1 | kaha\_1-wA\_hE\_1 |

Annotation:

* Here, *unheṃ* is a resumptive pronoun referring to *parvata*.
* It is **not** represented in the USR annotation.

## 

###### **V. Correlative Clauses**

###### **1.1 Characteristics of Correlative Clauses**

Correlative relative clauses differ from noun-modifying relative clauses. They:

* Occur at the **left periphery** of the main clause.
* Are headed by a **relative pronoun**.
* Must have a **correlate** (a demonstrative or pronominal) in the main clause.

###### **1.2 Annotation Strategy for Correlative Clauses**

1. The **finite verb** of the correlative clause is dependent on the **demonstrative/pronoun** in the main clause.
2. The **demonstrative pronoun or noun** in the main clause gets its **coreference** from the finite verb of the relative clause.

Example:

| Sent\_ID\_1 | jo laḍakā kala ghara āyā vaha merā bhāī hai | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | $yax | laḍakā\_1 | kala\_1 | ghara\_1 | ā\_1-yā\_1 | $wyax | $speaker | bhāī\_1 | hai\_1-pres |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Dep. Rel | 2:dem | 5:k1 | 5:k7t | 5:k2p | **6:rcdelim** | 9:k1 | 8:rhh | 9:k1s | 0:main |
| Discourse |  |  |  |  |  | **5:coref** |  |  |  |

| Sent\_ID\_2 | **#jaba rāma ghara jā rahā thā taba bāriśa ho rahī thī**  ‘When Rama was going home it was raining.’ | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | $yax | rAma | [ne\_1] | Gara\_1 | jA\_1-rahA\_WA\_1 | $wyax | bAriSa\_1 | ho\_1-rahA\_ WA\_1 | [cp\_1] |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| dependency | 5:k7t |  | 5:k1 | 5:k2p | **6:rcsamAnakAla** | 9:k7t |  |  | 0:main |
| Discourse element |  |  |  |  |  | **5:coref** |  |  |  |

**Annotation:**

* The relative clause *jo laḍa़kā kala ghara āyā* is linked to *vaha* in the main clause.
* The coreference is marked accordingly.

# **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 this 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.

## **Pronominal Coreference**

A discourse strategy to indicate two entities within a sentence or across sentences having the same referent.

#### **How to Annotate Pronominal Coreference**

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

| **Sent\_1** | rāma pustaka paḍha rahā hai | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| concept | rāma | [ne\_1] | pustaka\_1 | paḍha\_1-0\_ rahā\_hai\_1 | | |
| Index | 1 | 2 | 3 | 4 | | |
| Discourse element |  |  |  |  | | |
| **Sent**\_2 | **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**\_**1.2:coref** |  |  |  |  |  |

## **DISCOURSE CONNECTIVE**

Since USR annotation of complex sentences is difficult and automated USR generation for complex sentences is a challenge, complex sentences are first segmented into discourse units without losing information. The [**Discourse Connective Tag**](#_2nuputfaxut6) ensures that even after segmentation, the connective information is retained.

### **Annotation Strategy for Discourse Connective**

#### **Case 1: Single Connective in a Complex Sentence**

* Complex sentences are segmented into two simple sentences, with one containing the connective at the sentence level.
* The discourse relation tag is annotated on the main verb of the second sentence.
* The format: **Sent\_ID.Verb\_Index:Relation\_Name**
* The [connective](#_2nuputfaxut6) is **not** included in the USR.

| **Sent**\_1 | ***#***rāma skūla nahīṃ gayā kyoṃki vaha bīmāra hai  ‘Ram did not go to school because he is sick.’ | | | |
| --- | --- | --- | --- | --- |
| **Sent**\_1a | ***#***rāma skūla nahīṃ gayā  ‘Ram did not go to school.’ | | | |
| **concept** | ramā | skūla\_1 | nahīṃ\_1 | jā\_1-yā\_1 |
| index | 1 | 3 | 4 | 5 |
|  |  |  |  |  |
| **Sent**\_1b | ***#***kyoṃki vaha bīmāra hai  ‘Because he is sick.’ | | | |
| concept | $wyax | bīmāra\_1 | hai\_1-pres |  |
| index | 1 | 2 | 3 |  |
| Discourse element | Sent\_1a.2: coref |  | **Sent\_1a.4:kAryakAraNa** |  |

Table 1. Single Connective in complex sentence

#### **Case 2: Paired Connectives with a Main and Subordinate Clause**

* The main finite verb of the subordinate clause is chosen for discourse relation annotation.
* The connective is **not** included in the USR.
* The discourse relation tag conveys the relation between the two segmented sentences.

| **Sent**\_2 | ***#***yadi āpa mujhe āmaṃtrita karate haiṃ to maiṃ āpake ghara āūṃgā | | | |
| --- | --- | --- | --- | --- |
| **Sent**\_2a | ***#*** āpa mujhe āmaṃtrita karate haiṃ | | | |
| concept | $addressee | $speaker | āmaṃtrita+kara\_1-tā\_hai\_1 | |
| index | 1 | 2 | 3 | |
| Discourse element |  |  | **Sent**\_**2b.4:AvaSyakawApariNAma** | |
| **Sent**\_2b | ***#*** to maiṃ āpake ghara āūṃgā | | | |
| concept | speaker | addressee | ghara\_1 | ā\_1-gā \_1 |
| index | 1 | 2 | 3 | 4 |
| Discourse element |  |  |  |  |

Table 2. Paired connective

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

| **Sent**\_3 | yadi āpa acchā khānā khāoge aura āpa vyāyāma karoge to āpa svastha rahoge aura āpa bīmāra nahīṃ hooge | | | |
| --- | --- | --- | --- | --- |
| **Sent**\_3a | āpa ācchā khānā khāoge | | | |
| concept | $addressee | ācchā\_1 | khānā\_1 | khā\_1-gā\_1 |
| index | 1 | 2 | 3 | 4 |
| Discourse element |  |  |  |  |
| **Sent**\_3b | aura āpa vyāyāma karoge | | | |
| concept | $addressee | vyāyāma+kara\_1-gā\_1 |  |  |
| index | 1 | 2 |  |  |
| Discourse element |  | **Sent**\_**3a.4:samuccaya**  **Sent**\_**3d.4:AvaSyakawApariNAma** |  |  |
| **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**\_**3c.3:samuccaya** |

Table 3. Paired connective with coordination and subordination

# **CONSTRUCTION**

Semantic frames, which are linguistically expressed as *larger-than-lexical-but-smaller-than-sentential expressions*, are considered as Constructions (CxN) in Universal Semantic Representation.

The conceptual schema of the constructions are mentioned below:

(i) The CxN Frame as the Complex Concept

(ii) The semantic category of the components of the CxN.

The characteristic features of the Constructions are the following:

* There is a form or complex concept that maps to a meaning
* Each complex concept is given a name that reflects the meaning.
* Complex concepts are specified within bracket []
* The semantic tag of each slot of the pattern is specified in the Semantic Category feature row.
* The relations among the components of the construction are specified in the construction row.

### **How are complex concepts represented?**

The format : [name of construction\_number-of-occurence]

rāma aura sītā basa sṭapa para dūdha aura cāya pāna kiyā

‘Ram and Sita drank milk and tea at the bus stop.’

| Original Sentence | rāma aura sītā basa sṭapa para dūdha aura cāya pāna kiyā | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | rāma | sītā | [conj\_1] | basa\_1 | sṭapa\_1 | [6-waw\_1] | dūdha\_1 | cāya\_1 | [conj\_2] | pāna\_1 | kara\_1-yā\_1 | [cp\_1] | [ne\_1] | [ne\_2] |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| Semantic Category | male | female |  |  |  |  |  |  |  |  |  |  | per | per |
| Dependency relation |  |  | 12:k1 |  |  | 12:k7p |  |  | 12:k2 |  |  | 0:main |  |  |
| Construction relation | 13:begin | 14:begin |  | 6:mod | 6:head |  | 9:op1 | 9:op2 |  | 12:kriyAmUla | 12:verbalizer |  | 3:op1 | 3:op2 |

* *rāma* and *sītā* are two named entities, thus, they are components of named entity construction, and *rāma* and *sītā*  are *begin* names of these two named entities.
* Two named entities are intra-sententially conjoined, thus, they are the components of a conjunction construction. Conjunction construction is represented as **[conj\_1]** and *[ne\_1]* and *[ne\_2]*  are two operators **(op1 and op2)** of [conj\_1].
* There exists another intra-sentential conjunction in this sentence, *dūdha aura cāya* ‘milk and tea’. Thus, the complex concept of this construction gets **[conj\_2]**. *dūdha* and *cāya* are two operators of [conj\_2].
* *basa sṭapa* ‘bus stop’ is a noun compound, hence, gets the construction name with appropriate [noun compound (samasa) analysis](?tab=t.0#heading=h.3iugesx94u6), i.e.- **[6-waw\_1]** here. Components of a noun compound have internal modifier-modified relation here, thus, they get mod and head relation in respect to the construction frame, i.e.- The complex concept.
* pāna kiyā is a complex predicate. The construction name is identified as **[cp\_1]** for the [complex predicate](#_rp4qv0grd09e) with the **kriyAmUla** and **verbalizer** relations.
* The dependency relation of the construction is given to the complex concept. Thus, [cp\_1] gets 0:main and [conj\_1] gets k1, [conj\_2] gets k2 respectively.

### **Named Entity Construction**

A **Named Entity (NE)** is a word or phrase that refers to a specific entity, such as a person, organization, location, date, or product.

* Named entities introduce a complex structured concept, represented as **[ne\_1]**.
* Each entity is systematically tagged to ensure proper identification.
* **begin:** Used to mark the first word in a named entity (e.g., the first name in a person’s full name).
* **inside:** Applied to subsequent words within the same entity (e.g., middle name, last name, or additional words in an organization name).

| Original sentence | rāma ne cāya pī  ‘Rama drank tea.’ | | | |
| --- | --- | --- | --- | --- |
| concept | rāma | cāya\_1 | pī\_1-yā\_1 | **[ne\_1]** |
| Index | 1 | 2 | 3 | 4 |
| Sem. Cat. | male |  |  | **per** |
| Dep. Rel. |  | 3:k2 | 0:main | 3:k1 |
| Construction | **4:begin** |  |  |  |

### 

### **Compound Construction**

### **I. Noun Compound**

* Noun Compounds consist of two or more nouns with the final noun head and other noun modifiers or two nouns in coordinating relation as two operators.
* Noun Compounds can be spaced or hyphenated.
  + [[kṛṣi prasāraṇa] udyoga]
  + gṛha-śikṣaka
* When an adjective modifies the modifier, that adjective will be part of the noun compounds
  + [[vanyā jīva] surakṣā]
* One proper noun and one or more than nominal is NC
  + dillī pulisa

##### **List of Noun compound**

| Noun Compound Type | Example |
| --- | --- |
| karmaXAraya | nila kamala |
| wawpurusha | 2-waw **dhana prāpta** |
| 3-waw  **cintā grasta** |
| 4-waw  **guru dakṣiṇā** |
| 5-waw  **śoka mukta** |
| 6-waw **dhūla-kaṇoṃ** |
| 7-waw **viśva vitaraṇa** |
| xvanxa | **peḍa-paudhoṃ** |
| vahuvrIhi | **[tāpa-abhāva**] kṣetra |

###### **When NOT to consider as Noun Compound**

* We do not consider a noun compound as a construction when it is written together as ***vidyālaya*** ‘school’**.**
* **We do not consider**  a noun compound as a construction when **any post-position or suffix is posited in between**
  + lohe **kā** ciyara
* Compound Verb, i.e.- Verb + Light Verb
  + **kara ḍāl-ā [[1]](#footnote-0) ‘do keep-past’** 
    - rām ne sārā kāma ekadina maiṃ *kara ḍālā*

‘Ram could do the whole work in one day’

### **Noun Compound Construction**

[Noun Compound](#_3y3p0nm5uuv5) section explains what qualifies as a noun compound and what does not along with its types. We will demonstrate how to annotate noun compounds within a sentence.

* Noun compound construction is represented as [nc\_1] as an umbrella term, and the components are either *op1* and *op2* or *mod* and *head,* depending on the type of the compound.
* The generic [nc\_1] label can be replaced with more specific types of noun compounds, such as [6-waw\_1] or [xvanxva\_1].
  + **[6-waw\_1]**: [basa sṭapa]
  + **[xvanxva\_1]**: [candra-sūrya]
* The construction of noun compounds varies across languages. For instance, the same expression may take the form of an adjective-noun combination in one language, while in another, it appears as a noun-noun combination, where the first noun modifies the second. Example-
  + **Hindi**: *pahāḍaī ghoḍā* (adjective-noun combination)
  + **English**: *mountain horse* (noun-noun combination)

| Original Sentence | pitā-putra ne basa sṭapa para cāya pīyā  ‘Father and son had tea at the bus-stop.’ | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | pitā\_1 | putra\_1 | **[xvanxva\_1]** | basa\_1 | sṭapa\_1 | **[6-waw\_1]** | cāya\_1 | pī\_1-yā\_1 |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Semantic Category | anim/male | anim/male |  |  |  |  |  |  |
| Dependency relation |  |  | 8:k1 |  |  | 8:k7p | 8:k2 | 0:main |
| Construction relation | **3:op1** | **3:op2** |  | **6:mod** | **6:head** |  |  |  |

### **II. Compound Construction**

* When the head of the compound **is not a noun,** but the **modifier is a noun**, we name the construction as [compound\_1]
  + [namaka yukta]
  + **[bhūkaṃpa saṃbaṃdhī]** kriyāeṁ

### 

| Original Sent | anya apratyakṣa srotoṃ meṃ bhūkaṃpa saṃbaṃdhī kriyā śāmila haiṃ | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | anya\_1 | apratyakṣa\_1 | srota\_1 | bhūkaṃpa\_1 | saṃbaṃdhī \_1 | **[compound\_1]** | kriyā\_1 | śāmila\_1 | hai\_1-pres |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Sem Cat |  |  |  |  |  |  |  |  |  |
| Dep Rel | 2:mod | 3:mod | 7:rn |  |  | 7:mod | 9:k1 | 9:k1s | 0:main |
| Construction |  |  |  | **6:mod** | **6:head** |  |  |  |  |

### 

### **Complex Predicate**

A complex predicate consists of a kriyāmūla and a kriyā or verbalizer. They are represented as two different concepts with specific concept ID and the complex concept [cp\_1] denotes their complex construction.

rāma ne  **snāna kiyā**.

‘Rama took a bath.’

| Original Sentence | rāma ne snāna kiye. ‘Rama took bath.’ | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | ramā | [ne\_1] | snāna\_1 | | | | kara\_1-yā\_1 | **[cp\_1]** |
| Index | 1 | 2 | 3 | | | | 4 | 5 |
| Sem. Cat. | male | per |  | | | |  |  |
| Dependency row |  | 5:k1 |  | | | |  | 0:main |
| Construction | 2:begin |  | **5:kriyAmUla** | | | | **5:verbalizer** |  |

### **Conjunction Construction**

* A conjunction construction consists of more than one element which are conjoined together and are in identical dependency relation with the head.
* Components of [conj\_1] constructions have the tag of op1, op2…

index of [conj\_1]:op

Conjunction Construction

| Original Sentence | rāma aura mohana skūla gaye.  ‘Rama and Mohana went to school.’ | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | ramā | mohanā | | **[conj\_1]** | skūla\_1 | jā\_1-yā\_1 | | | | [ne\_1] | [ne\_2] |
| Index | 1 | 2 | | 3 | 4 | 5 | | | | 6 | 7 |
| Sem. Cat. | male | male | |  |  |  | | | | per | per |
| Dependency row |  |  | | 5:k1 | 5:k2p | 0:main | | | |  |  |
| Construction Row | 6:begin | 7:begin | |  |  |  | | | | **3:op1** | **3:op2** |

### **Disjunction Construction**

* A disjunction construction expresses the relation between more than one element which are distinct alternatives.
* Relation among the components are given in the construction row as following

Index of [disjunct\_1]:opn

Disjunction Construction

| Original Sentence | rāma roṭī yā cāvala khāyegā ‘Rama will eat bread or rice.’ | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | rām | roṭī\_1 | | cāvala\_1 | **[disjunct\_1]** | khā\_1-gā\_1 | | | | [ne\_1] |
| Index | 1 | 2 | | 3 | 4 | 5 | | | | 6 |
| Sem. Cat. | male |  | |  |  |  | | | | per |
| Dependency row |  |  | |  | 5:k2 | 0:main | | | | 5:k1 |
| Construction | 6:begin | **4:op1** | | **4:op2** |  |  | | | |  |

### **Calendar Construction**

* A calendar construction is consists of following elements in an order

Year>Month>Date>Hour>Minute>Second

* Semantic Category row gives the semantic tag of each slot of the pattern.
* Relation among the components are given in the construction row as following

Index of [calendar\_1]:component

Calendar Construction

| Original Sentence | 15 julāī 2020 rāma paidā huā thā  ‘Ram was born on 15th July, 2020.’ | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | 15 | julāī\_1 | 2020 | [calendar\_1] | rAma | paidā\_1 | ho\_1-yā\_ thā\_1 | [cp-1] | [ne\_1] |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Sem. Cat. | dom | moy | yoc |  | male |  |  |  | per |
| Dependency row |  |  |  | 8:k7t |  |  |  | 0:main | 8:k1 |
| Construction | 4:component | 4:component | 4:component |  | 9:begin | 8:kriyAmUla | 8:verbalizer |  |  |

### **Spatial Construction**

* A spatial construction shows the relation as a part-whole between the places as

whole place> place within

* The relation is shown as

Index of [spatial\_1]: whole and Index of [spatial\_1]:part

| Original Sentence | bhārata meṃ meghālaya rājya ke cūne kī śailoṃ ke pradeśa meṃ raṃdhra dekhane ko milate haiṃ| | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | bhārata | meghālaya | rājya\_1 | **[spatial\_1]** | cūnā\_1 | śaila\_1 | pradeśa\_1 | raṃdhra\_1 | dekha\_1 | mila\_1-tā\_ hāi\_1 | [ne\_1] | [ne\_2] |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Sem. Cat. |  |  |  |  |  |  |  |  |  |  | place | place |
| Dependency row |  |  | 2:rs | 5:r6 | 6:r6 | 7:r6 | 10:k7p | 10:k1 | 10:rt | 0:main |  |  |
| Construction | 11:begin | 12:begin |  |  |  |  |  |  |  |  | **4:whole** | **4:part** |

## 

### **Temporal Construction**

* A temporal construction shows the relation as a part-whole between the times as

**whole time> time within**

* The relation is shown as

Index of [temporal\_1]: whole and Index of [temporal\_1]:part

| Original Sentence | manu subaha sāta baje calanā śurū kiyā |  ‘Manu started walking at 7 in the morning.’ | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | manu | subaha\_1 | 7 | **[temporal\_1]** | cala\_1 | śurū\_1 | kara\_1-yā \_1 | [cp\_1] | [ne\_1] |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Sem. Cat. | male |  | clocktime |  |  |  |  |  | per |
| Dependency row |  |  |  | 8:k7t | 8:k2 |  |  | 0:main | 8:k1 |
| Construction | 9:begin | **4:whole** | **4:part** |  |  | 8:kriyAmUla | 8:verbalizer |  |  |

## 

### **Span Construction**

* A span construction shows the relation as a span of the full extent or range of something from start to end.
* Name of this construction is [span\_1] and the components are **start and end**.
* The relation is shown as

Index of [span\_1]: start and Index of [span\_1]:end

| 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 | **[span\_1]** |
| Index | 1 | | 2 | | | 3 | | 4 | 5 |
| Sem. Cat. | yoc | | yoc | | |  | |  |  |
| Dependency row |  | |  | | | 4:k1 | | 0:main | 4:k7t |
| Construction | **5:start** | | **5:end** | | |  | |  |  |

### **Measurement Construction**

* Measurement is the process of quantifying the attributes of an object or event. The result of a measurement is a numeric value with certain units.
* Name of measurement construction is  **[x\_meas\_1]**, components are **count and unit.**
* The relation is shown as

Index of [x\_meas\_1]: count and Index of [x\_meas\_1]:unit [x=distance, length, height, temperature, weight, distance,width, depth, volume, speed]

| Original sentence | rāma 10 kimi cale  ‘Rama walked 10 kilometers.’ | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| concept | rāma | 10 | kilomitara\_1 | cala\_1-yā\_1 | [ne\_1] | **[dist\_meas\_1]** |
| Index | 1 | 2 | 3 | 4 | 5 | 6 |
| Sem. Cat. | male | numex |  |  | per |  |
| Dep. Rel. |  |  |  | 0:main | 4:k1 | 4:rmeas |
| Construction | 5:begin | **6:count** | **6:unit** |  |  |  |

### **Rate Construction**

* The concept of **rate** involves two measurements, often involving time, quantity, amount, or degree of something measured per unit of something else.
* Name of rate construction is  **[rate\_1]**, components are **unit\_value** and **unit\_every** consists of value/per unit.
* Value can be a measuring expression or a counting expression (cardinal+noun expression)
* Unit can be a measuring unit or measuring noun.

| Original sentence | rāma 10 kimi prati ghaṃṭā cale  ‘Rama walked 10 kilometers each hour.’ | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| concept | rāma | 10 | kilomitara\_1 | 1 | ghaṃṭā\_1 | **[rate\_1]** | cala\_1-yā\_1 | [ne\_1] |
| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Sem. Cat. | male | numex |  |  |  |  |  | per |
| Dep. Rel. |  |  |  |  |  | 7:rmeas | 0:main | 7:k1 |
| Construction | 8:begin | **6:count** | **6:unit\_value** | **6:count** | **6:unit\_every** |  |  |  |

### 

### 

### 

## **APPENDIX**

### **Segmentation**

##### **I. Discourse Units and Finite Verbs**:

Segmented discourse units typically contain a finite verb. A discourse unit is a simple sentence or clause, which may not always be the smallest unit but contributes to the overall discourse.

Example:

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

"Ram and Sita spoke to their brother at the bus stand."

##### **II. Relative Clauses**:

* **When Not to Segment**:  
  Relative clauses referring to a noun within the sentence are not segmented.  
  Example:
  + *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 2004."
* **When to Segment**:  
  Sentences with multiple relative clauses are segmented, with inter-clausal relations annotated in the discourse element row.
* Original Sentence:

| 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| |
| --- | --- |

* Original Sentence:

| 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| |
| --- | --- |

* Segmented Outcome

| 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 Segmentation](#_gierfav9ryf5) for detailed segmentation strategies.

**Relative Pronouns as Discourse Connectives**:

If a relative pronoun functions as a discourse connective, the relative clause is segmented.

Example:

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

* Original Sentence:

| Sent\_ID\_1 | *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.* |
| --- | --- |

* Segmented Outcome

| Sent\_ID\_1a | nadī ke nicale bhāgoṃ meṃ ḍhāla kama hone ke kāraṇa nadī kī gati kama ho jātī hai. |
| --- | --- |
| Sent\_ID\_1b | isake pariṇāmasvarūpa nadīya dvīpoṃ kā nirmāṇa hotā hai. |

##### **Relative Clause Segmentation**

##### Case A:

One head, one relative clause – won’t be segmented

**rāma**, [jo merā bhāī hai] saṃskṛta kā chātra hai

**rāma** saṃskṛta kā chātra hai , [jo merā bhāī hai]

[jo merā bhāī hai] **vaha** saṃskṛta kā chātra hai

##### Case B:

One head, multiple relative clauses – will be segmented – relative pronoun will be added in the segmented relative clause

**rāma** [jo merā bhāī hai] [aura govā meṃ rahatā hai] kala yahāṃ āyā hai

a. rāma kala yahāṃ āyā hai

b. jo merā bhāī hai

c. **jo** govā meṃ rahatā hai

[jo govā meṃ rahatā hai] aura [saṃskṛta paḍha़tā hai] **vaha** merā bhāī hai

a. jo govā meṃ rahatā hai

b. **jo** saṃskṛta paḍha़tā hai

c. vaha merā bhāī hai

##### Case C:

Two heads three relative clauses, two relative clauses are modifying one head, one relative clause is modifying one head – segmentation rule will be following Case I + Case II

[0.2] **rāma** ne [jo merā bhāī hai], [aura govā meṃ rahatā hai], **eka kitāba** likhī hai [jise kaī puraskāra mile haiṃ]

[0.3] **rāma** [jo merā bhāī hai], [aura govā meṃ rahatā hai] ne, **eka kitāba** likhī hai [jise kaī puraskāra mile haiṃ]

a. rāma ne eka kitāba likhī hai jise kaī puraskāra mile haiṃ (case I)

b. jo merā bhāī hai

c. **jo** govā meṃ rahatā hai (b and c following case II)

##### Case D:

One relative clause two main clauses, relative clause will not be segmented, main clause will be segmented and pronoun will be added in the segmented main clause

**rāma** [jo kala yahāṃ āyā hai] merā bhāī hai aura govā meṃ rahatā hai

a. rāma jo kala yahāṃ āyā hai merā bhāī hai

b. aura **vaha** govā meṃ rahatā hai,

But, in the case of a correlative clause, if there are two main clauses and both of them are coreferred with the relative clause, then the relative clause itself will be segmented along with the main clauses.

[jo merā bhāī hai] **vaha** govā meṃ rahatā hai aura saṃskṛta paḍha़tā hai

a. jo merā bhāī hai

b. vaha govā meṃ rahatā hai

c. aura **vaha** saṃskṛta paḍha़tā hai

##### III. **Complement Clause Segmentation**

* + The word *yaha* ("this") is added to the main clause to co-refer to the complement clause.

Example:

* Original Sentence

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

* Segmented Outcome

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

* **Special Case: itanā...ki as a Discourse Connective**:

When *itanā...ki* functions as a discourse connective, sentences are segmented, and *isase* ("this") is postulated as the connective which brings ‘pariNama’ relation and add iwanA\_ki in the speaker's view row.

Example:

* 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| |
| --- | --- |

* Segmented outcome

| 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| |

##### **IV. Connectives**

1. **Single Connectives**:  
   When two clauses are connected by a single connective, they are segmented, retaining the connective in its original clause.  
   Example:

* Original Sentence

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

* Segmented outcome

| 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.’ |
| --- | --- |

* Segmented outcome

| 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. |
| --- | --- |

* Segmented outcome

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

**2. Paired Connectives**:

Paired connectives are segmented into two sentences, retaining the connective in the main clause.

Example:

* 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’ |

By employing these strategies, the USR annotation ensures detailed and accurate representation of discourse connectives and relations without losing critical information.

### 

### 

### **Concepts**

##### **I. List of Complex Concepts**:

| Complex Concept Type | Complex Concept Representation | Example |
| --- | --- | --- |
| [Named Entity](#_yy2gbohmzhqj) | [ne\_1] | **Atala Bihari Vajpayee** |
| [Noun Compound](#_3y3p0nm5uuv5) | [nc\_1] | **ūrjā vikiraṇa** |
| [Complex Predicate](#_rp4qv0grd09e) | [cp\_1] | **snāna kiyā** |
| [Measuring expression](#_wtpbfw3bcwqy) | [time\_meas\_1] | **10 ghaṃṭe** |
| [dist\_meas\_1] | **10 kimi\*** |
| [length\_meas\_1] | **6 phīṭa** |
| [width\_meas\_1] | **6 phīṭa** |
| [temp\_meas\_1] | **5 ḍigrī+selsiyasa** |
| [depth\_meas\_1] | **6 phīṭa** |
| [height\_meas\_1] | **6 phīṭa** |
| [volume\_meas\_1] | **100 līṭara** |
| [weight\_meas\_1] | **10 kilo** |
| [Calendric expression](#_guaakp66nbm7) | [calendar\_1] | **5 julāī 2020** |
| [Spatial expression](#_wy1hbof5okcd) | [spatial\_1] | **bhārata meṃ meghālaya rājya** |
| [Rate](#_jnnlw2r2gyh8) | [rate\_1] | **80 kimī prati ghaṃṭā , prati do ghaṃṭe meṃ eka bāra** |
| [Span](#_fe4iraqtub7w) | [span\_1] | **1990 se lekara 2000 taka, aba taka** |
| [Conjunction](#_7bu50mor561r) (intra-sentential) | [conj\_1] | **rāma aura mohana** |
| [Disjunction](#_3pfusscq91b5) (intra-sentential) | [disjunct\_1] | **roṭī yā kelā** |

\***S.I. units, such as-kimi, kilo will be written as full form- kilomitara, kilogrāma etc**

##### **II. Which items do not get Concept ID in the concept row**

| Name of Concept | Example |
| --- | --- |
| Named Entity | **rāma, bhārata** |
| Pronominal | **$speaker, $addressee, $wyax,$yax** |
| Numeral expression (year, date or cardinal) | **10, 1990** |
| Abbreviation | **@eic.sl.yU., @nAsA (NE)** |

##### 

##### **III. Which items are not presented in the concept row**

| Name of Category | Example |
| --- | --- |
| Connectives | aura ‘and’, yā ‘or’, kyoṃki ‘because’, isaliye ‘therefore’, udāharansvarupa ‘for example’, jese ‘as such’, yāthā ‘as an example’ |
| Discourse Particle | hī, bhī, sirpha,khāsakara, visheshkara |
| 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?’ |
| pratyeka/ prati as a part of a rate construction/ in-every/ in-each | prārambha meṃ tāpamāna baDne kī ausata dara **pratyeka** *32 mīṭara kī girāī para 1⁰ selsiyasa* hai| |

##### 

##### **IV. Concept representation of all Pronouns**

| Personal Pronoun | 1st Person | *# tuma mere ghara āo*  ‘You come to my home.’ | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concept | $addressee | **$speaker** | | | ghara\_1 | | | ā\_1-o\_1 | | | | | |
| 2nd Person | *# tuma mere ghara āo*  *‘*You come to my home.’ | | | | | | | | | | | | |
| Concept | **$addressee** | $speaker | | | ghara\_1 | | | ā\_1-o\_1 | | | | | |
| 3rd Person Pronominal | | #usane nahīṃ khāyā.  ‘He did not eat.’ | | | | | | | | | | | | |
| Concept | | **$wyax** | nahīṃ\_1 | | | | | | khā\_1-yā\_1 | | | | | |
| Reciprocal Pronoun | | *#hama eka dūsare se pyāra karate haiṃ.*  *‘*We love each other.’ | | | | | | | | | | | | |
| Concept | | $speaker | **eka+dūsarā** | | | | | | pyāra\_1 | | kara\_1-tā\_hā\_1 | | [cp\_1] | |
| Reflexive Pronoun | | #unhoṃne apanā kāma khuda kiyā  ‘They themself did their work’ | | | | | | | | | | | | |
| Concept | | $wyax | **apanā** | | kāma | | **khuda** | | kara\_1-yā\_1 | | | | | |
| Interrogative Pronoun | | #rāma kyā khā rahā hai? | | | | | | | | | | | | |
| Concept | | rāma | **$kim** | | | | | | khā\_1-0\_rahā\_hai\_1 | | | | | |
| Relative Pronoun | | #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 the water in the year 2004. | | | | | | | | | | | | |
| Concept | | bhārata | dakṣiṇī\_1 | biṃdu\_1 | $yax | iṃdirā+biṃdu | kaha\_1-yā\_jātā\_thā\_1 | san+2004 | jalamagna\_1 | | | ho\_1-gayā\_1 | | |

##### **V.** **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 |

##### **VI. 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 | Spatial near |  |

##### **VII. Convention of symbols used in USR**

| Concept | Symbol | Explanation | Example |
| --- | --- | --- | --- |
| 1st, 2nd and 3rd person pronouns | $ | All pronominals are flagged with ‘$’ and they do not get concept ID. | $speaker, $addressee, $wyax, $yax, $kim, |
| Foreign word | ^ | 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. | ^word\_1 |
| Abbreviation | @ | 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. | @eic.sl.yU., @nAsA |

# 

##### VIII. List of Quantifiers in Hindi

###### 

| saba |  |
| --- | --- |
| kuCa |  |
| koi |  |
| prawyeka |  |
| WodZA |  |
| saBI |  |
| kaI |  |
| kisi\* |  |
| hara |  |
| prawi |  |

\*kisi will be written as koi in concept

### 

### **Speaker’s View**

##### **I. List of Discourse Particle**

### 

| Meaning | Tag | Example |
| --- | --- | --- |
| Distinction, distinguishing, exclusive | hI\_1 | mujhako patā hai rāma **hī** jāyegā  ‘I know that Ram (and no one else) will go.’ |
| fixture,firmness | hI\_2 | merī do bahanoṃ kī **hī** śādī ho cukī hai  ‘Both my sisters got married.’ |
| A few | hI\_3 | kucha **hī** dinoṃ meṃ sardī calī jāegī  ‘The cold will go away in a few days.’ |
| Right from | hI\_4 | suru se **hī** isa talāo ke pānī meṃ namaka jyādā hai  ‘There is too much salt in this pond’s water from the very beginning.’ |
| Not only…but also | hI\_5 | kevala mohana **hī** nahī para raunaka bhī isa kāma meṃ juṭe hue hai  ‘Not only Mohan, but Rounak is also engaged in this work.’ |
| Only | hI\_6 | bacā huā kāma pūrā karane meṃ 10 **hī** dina lageṃge  ‘Only 10 days would be required to finish the remaining work.’ |
| Additional (but not as conjunctive) | aura\_1 | eka **aura** cāyale āo  ‘Bring one more tea.’ |
| also/Inclusive | BI\_1 | sūrya **bhī** eka tārā hai.  ‘The sun is a star.’ |
| Emphasis | BI\_2 | ye pavana 36 ghaṃṭe se **bhī** j​yādā samaya taka calatī rahī.  ‘This wind lasted for over 36 hours.’ |
| Any | BI\_3 | ārdra evaṃ garma mausama **kisī ko bhī** cidhcidhā banā sakatā hai ।  ‘A hot or humid weather may make one irritable. |
| Yet/Even then | BI\_4 | hālāṁki kāphī rāta ho cukī thī **phira bhī** jonaka ne jaṃgala meṃ praveśa karane kā phaisalā kiyā  ‘ALthough it was late at night yet Jonak decided to enter the forest.’ |
| Still | BI\_5 | vaijñānika **abhī bhī** blaika holake bāre meṃ adhika se adhika jānakārī ekatra karane meṃ juṭe haiṃ.  ‘Scientists are still trying to find out more and more about the Black Hole.’ |
| only | kevala\_1 | sammelana meṃ **kevala** 10 loga śāmila hue  ‘Only 10 people attended the conference.’ |
| Modifier of adjective/ intensifier | sA\_1 | mohana ne apanī beṭī ke lie eka choṭā **sā** khilaunā pillā kharīdā  ‘Mohan bought a small toy puppy for his daughter.’ |
| turn-taker | wo\_1 | **to**, āpa ina dinoṃ kyā kara rahe haiṃ?  ‘So, what are you upto these days?’ |
| Nearly | kariba\_1 | mahākuṃbha mele meṃ **karība** 4 karoḍa śraddhālu juṭe haiṃ  ‘Nearly 4 crores devotees have gathered at Mahakumbh Mela.’ |
| Almost/ approximate | lagaBAga\_1 | ṭrena lagabhaga 12 ghaṃṭe kī derī se cala rahī hai  ‘The train is running almost 12 hours.’ |
| Only | sirPa\_1 | abhī taka **sirpha** 30 phīsadī kāma hī pūrā huā hai  ‘Only 30% of work has been finished so far.’ |
| Affirmation | hAz\_1 | hāṁ, hama cuniṃdā deśoṃ meṃ aṃtarrāṣṭrīya śipiṃga pradāna karate haiṃ|  ‘Yes, we offer international shipping to select countries.’ |

##### **II. 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 |  |
| ā\_1 |  |

### **Discourse Connective**

##### I. **List of Discourse Connectives**

| **Name of discourse relation** | **Marker** | **Tag** | **Example** | **Explanation** |
| --- | --- | --- | --- | --- |
| āvaśyakatā 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. |
| āvaśyakatā pariṇāma.nahīm | nahīm…to/ agara…to/ to | AvaSyakawApariNAma.nahIM | rāma āegā  nahīm to maiṃ jāūṃgī.  ‘If Ram does not come then I will go.’ | This relation presents a condition where one action will occur if a specified condition is met, and an alternative outcome will happen if the condition is not fulfilled. |
| anyatara | 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. | This relation explains the reason or basis for an event or action, showing why something happens. It connects a specific cause to its resulting effect. |
| pariṇāma | 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.’ | This relation connects two parts of a sentence where one part (the cause) leads to or produces a specific outcome or consequence (the result). |
| 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. |
| samuccaya.atirikta | isake atirikta | samuccaya | rāma āḍaī 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.alāvā | 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 | | When additional information is added to an existing one, stated before, we use the samuccaya relation. **alAvA** information encoded in the speaker's view row. |
| samuccaya.bhī | na kevala..balki | samuccaya | **nā kevala** rāma paḍha़āī meṃ acchā hai **balki** khela kūda meṃ bhī bahuta āge hai| | This relation connects two related actions or ideas, where the first part introduces one action or fact, and the second part adds another action or fact that complements or reinforces the first.  We use **samuccaya** relations. **BI** information encoded in the speaker's view row. |
| samuccaya.samāveśī | isake sāth sāth | samuccaya | rāma khānā banātī hai| **isake sātha-sātha**, vaha ghara kā kāma bhī karatī hai| | When additional information is added to an existing one, stated before, **samuccaya** is the discourse element tag and **samAveSI** is the speaker’s view tag |
| 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. |
| virodhī\_dyotaka | jabaki | viroXI\_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 | tāki | 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 | | This relation connects an action to its intended goal or objective. It shows that one action is done with the intention of achieving a specific result or purpose. |
| arthāt | dūsare śabdoṃ meṃ/arWAwa | arWAw | 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. |
| udāharaṇasvarūpa | 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. |

### Dependency Relation

##### K1

| Original sentence | rāma āma khātā hai| | | | |
| --- | --- | --- | --- | --- |
| concept | rāma | [ne\_1] | āma\_1 | khā\_1- tā\_hai\_1 |
| Index | 1 | 2 | 3 | 4 |
| Dep. Rel. |  | **4:k1** | 4:k2 | 0:main |

K2

| Original sentence | mohana ne āma kharīde| | | | |
| --- | --- | --- | --- | --- |
| concept | mohan | [ne\_1] | **āma\_1** | kharīda\_1-yā\_1 |
| Index | 1 | 2 | 3 | 4 |
| Dep. Rel. |  | 4:k1 | **4:k2** | 0:main |

K3

| Original sentence | mohana ne āma cākū se kāṭā| | | | | |
| --- | --- | --- | --- | --- | --- |
| concept | mohan | [ne\_1] | āma\_1 | cākū\_1 | kāta\_1-yā\_1 |
| Index | 1 | 2 | 3 | 4 | 5 |
| Dep. Rel. |  | 5:k1 | 5:k2 | **5:k3** | 0:main |

K4

| Original sentence | bhāī ne **billī ko** dūdha dīyā | | | |
| --- | --- | --- | --- | --- |
| concept | bhāī | **billī\_1** | dūdha \_1 | dī\_1-yā\_1 |
| Index | 1 | 2 | 3 | 4 |
| Dep. Rel. | 4:k1 | **4:k4** | 4:k2 | 0:main |

K5

| Original sentence | **peḍa se** eka pattā girā | | | |
| --- | --- | --- | --- | --- |
| concept | **peḍa\_1** | eka\_2 | pattā\_1 | gira\_1-yā\_1 |
| Index | 1 | 2 | 3 | 4 |
| Dep. Rel. | **4:k5** | 3:quant | 4:k1 | 0:main |

K7

| Original sentence | mohana **rājanīti** **para** carcā kara rahe the | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| concept | mohana | **rājanīti**\_1 | carcā\_1 | kara\_1-0\_rahā\_thā\_1 | [cp\_1] | [ne\_1] |
| Index | 1 | 2 | 3 | 4 | 5 | 6 |
| Dep. Rel. |  | 5:k7 |  |  | 0:main | 5:k1 |
| Cxn | 6:begin |  | 5:kriyāmūla | 5:verbalizer |  |  |

K7t

| Original sentence | rāma **cāra baje** āyegā | | | |
| --- | --- | --- | --- | --- |
| concept | rāma | [ne\_1] | **4** | āyā\_1-gā\_1 |
| Index | 1 | 2 | 3 | 4 |
| Semcat |  |  | clocktime |  |
| Dep. Rel. |  | 4:k1 | **4:k7t** | 0:main |

K7p

| Original sentence | **meja** para kitāba gira gayā hai | | |
| --- | --- | --- | --- |
| concept | meja | kitāba\_1 | gira\_1-yā\_hai\_1 |
| Index | 1 | 2 | 3 |
| Dep. Rel. | **3:k7p** | 3:k1 | 0:main |

K4a

| Original sentence | **rāma ko** āma pasaṃda hai | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| concept | **rāma** | **[ne\_1]** | āma\_1 | pasaṃda\_1 | hai\_1-pres | [cp\_1] |
| Index | 1 | 2 | 3 | 4 | 5 | 6 |
| Dep. Rel. |  | 6:k4a | 6:k1 |  |  | 0:main |
| Cxn | 2:begin |  |  | 6:kriyāmūla | 6:verbalizer |  |

1. Check raMjak kriya section for annotation of compound verbs [↑](#footnote-ref-0)