USR Guidelines

Universal Semantic Representation (USR) captures the meaning expressed by a sentence in the discourse. The meaning is represented in 10 rows in csv (comma (,) separated value) format. These rows are:

1. Original Sentence
2. Local word grouping
3. Concept representing content words in the sentence and TAM (tense-aspect-modality) specification on the verb
4. Index for the groups
5. Semantic Category of nouns
6. GNP (Gender, Number, Person) Information
7. Intrachunk Dependency Relations
8. Interchunk Dependency Relations
9. Extra-propositional meaning
    a. Negation
    b. Modality
    c. Attitude
    d. Discourse particles
10. Discourse elements:
    a. Anaphora
    b. Connectives
11. Sentence Types

We will explain the details of each row with the intention of generating the following sentence from the information represented in the rows of the USR.

"राम ने बस अड्डे पर एक अच्छे लड़के के साथ ही बात की":

Row 1: Sentence to be generated

Row 1:राम ने बस अड्डे पर एक अच्छे लड़के के साथ ही बात की .

The sentence represents the thoughts that are conveyed in the following 9 rows, The sentence can be written in roman script or in original script such as in devanagari script for Hindi.

Row 2 : Local Word Grouping

- Separate the local word groups with comma

Local word group involves two cases:

- Noun (with its post-position, if any) and its modifiers
- Verb with its TAM
- Negation
- Discourse particles

Local word grouping for our example sentence is the following:

राम ने, बस अड्डे पर, एक अच्छे लड़के के साथ, ही, बात की, .

Adjectives, Quantifier and संख्या/number are grouped with nouns:

Ex. एक अच्छे लड़के के साथ

Adjectives – अच्छा लड़का, छोटा लड़का

Quantifier – बहुत काम

Adverbs – धीरे

Other words such as 'साहब', 'जी' are also considered as particles and grouped with the noun.

saMKyA (number) – A number can be followed by a measuring unit or by noun or (viSeRaNa+noun).

Example of the following word being a measuring unit :

चार किलोमीटर, 4 किलोमीटर

Example of the following word being a noun or (viSeRaNa+noun):

Ex. 4 लड़के, चार लड़के, (चार (छोटे लड़के))

परिमाणवाची (Measurement words) – किलोमीटर, मीटर, ग्राम

A measuring unit (see the list), can be followed by a noun or (viSeRaNa+noun):

Ex. (((चार किलोमीटर) लम्बा) रस्ता), ((एक दर्जन) केला)

क्रमवाची (ordinal numbers) – प्रथम, द्वितीय, पहला, दूसरा

क्रमवाची words can precede a noun or (viSeRaNa+noun):

Ex. पहला घर, (पहला (जातीय विद्यालय))

Negation: A negation marker will always be written as a separate group before or after the verb group if it occurs inside the verb group with which it occurs:

Row–1: राम पुस्तक नहीं पढ रहा है.

Row–2: राम, पुस्तक, नहीं, पढ रहा है,.

Row–1: राम पुस्तक पढ नहीं रहा है.

Row–2: राम, पुस्तक, नहीं, पढ रहा है.

Discourse Particles are grouped separately – तक, ही, भी

नहीं, राम पुस्तक पढ रहा है.

Here, नहीं is a discourse particle

Comparatives – तर, तम, सबसे

Comparatives are joined with the adjectives:

Ex. सबसे अच्छा, सुन्दर तम, सुन्दर तर

Connectives are also grouped separately. Some connectives are कि, आदि, और and so on.

End of the sentence – The punctuation occurs at the last column of row 2.

Row 1: #राम ने बस अड्डे पर एक अच्छे लड़के के साथ ही बात की .

Row 2: #राम ने,बस अड्डे पर,एक अच्छे लड़के के साथ, ही,बात की,.

Row 3: Concept from Concept dictionary

This row represents the concepts. Each concept is formally represented by one or more words (Multiword Expression (MWE)) in the generated sentence. These concepts are listed in the ***concept dictionary*** with a unique id. For example, the concept ids A_1 and A_2 in the following table represent two different concepts as can be seen in the English column:

Sense_Label	**Hindi_L**	**Eng_L**	**Example**
A_1	A_1	come_1	rAma Gara AwA hE "Ram comes home"
A_2	A_2	know_1	rAma ko hindI AwI hE "Ram knows Hindi"
Pala_1	Pala_1	fruit_1	muJe eka Pala xo "Give me a fruit"
Pala_2	Pala_5	result_2	bure kAma bure Pala xewe haiM. "Bad deeds give bad results."

Why are the words in Hindi and English columns also indexed? It is because they are the concepts (and not words). From the above table, we understand that the Hindi concept A_1 is equivalent to the concept of *comingcom* in English (come_1) while Hindi A_2 corresponds to the English concept know_1.

Like content words, verb TAM markers can also be polysemous in nature and that is why they are also represented in TAM concept dictionary can be of tinganta (finite) form or kridanta (non-finite, verbal noun, participial) form.

The TAM is separated from the root by '-' (hyphen). When the root is in bare form, we postulate a zero.
The mulitword TAM is written with an underscore.
Following are the representation of

करेगा, कर रहा है, किया था

कर–गा_1, कर–0_रहा_है_1 कर–या_था_1

In Hindi the stative predicate is denoted by the word है in present tense and था in past tense. The stative predicate occurs in various context and accordingly we have classified them as follows:

a. **State_copula:**

   Row 1: .राम अच्छा है.

   Row 2 : राम, अच्छा,है,.

   Row 3: rAma,acCA_1,state_copula–pres

b. **State_existential**

   Row 1: .भगवान हैं.

   Row 2 : भगवान,है,.

   Row 3: BagavAna,state_existential–pres

c. **State_possession**

   Row 1: राम के पास किताब है

   Row 2: राम के पास, किताब, है,.

   Row 3: rAma,kiwAba,State_possession–pres

d. **state_Experientia**l

   Row 1: राम को बुखार है

   Row 2: राम को, बुखार, है

   Row 3: rAma,buKAra_1,state_experience–pres

e. **State_Location**

   Row 1: राम हैदराबाद में है

   Row 2: राम, हैदराबाद में, है

   Row 3: rAma,hExarAbAxa,state_location–pres

The verb–TAM info of for stative predicates is given below:

State_copula–pres

Particles can also be ambiguous and therefore they are also given concept ids. For example, the meaning of ही in the following two sentences are different:

राम **ही** जाएगा (implies : no one else will go)

अब तुम शादी कर **ही** लो (implies : you have waited for long…)

शुरु शुरु में सब कुछ अच्छा **ही** लगता है (implies: the situation can change after some time)

Note:

A. No concept id will be used for pronouns or proper nouns.
B. For the correct concept id, check the TAM dictionary, particle dictionary and concept dictionary.

Complex Predicate – A complex predicate consists of a क्रियामूल and a क्रिया. They are joined by "+".

For स्नान किया -> स्नान+कर_1–या_1

Compound noun – A compound noun is made up of a head noun and one or more noun modifiers. They are also joined with "+".

For बस अड्डा -> बस_1+अड्डा_1

Row 1: #राम ने बस अड्डे पर एक अच्छे लड़के के साथ ही बात की .

Row 2: #राम ने,बस अड्डा पर,एक अच्छा लड़का के साथ, ही,बात की,.

Row 3: राम,बस_1+अड्डा_1,एक_1 अच्छा_1 लड़का_1, ही_1,बात+कर_1–या_1

Row 4: Index for Groupings

Each concept (i.e. the prakrti) is indexed

Row 1: #राम ने बस अड्डे पर एक अच्छे लड़के के साथ ही बात की .

Row 2: #राम_ने,बस+अड्डा_पर,एक अच्छा लड़का_के_साथ, ही,बात+कर–या,.

Row 3: राम,बस_1+अड्डा_1,एक_1 अच्छा_1 लड़का_1, ही_1,बात+कर_1–या_1

Row 4: 1,2,3,4,5

One thing to be noted here is that each content word within a local word group has a sub-index-id. For example, in the above sentence, एक is 3.1, अच्छा is 3.2, लड़का is 3.3. We will utilize these sub-index-ids to express dependency relations in 7th, 8th and 9th rows.

Row 5: Semantic Categories of Nouns

Semantic category row captures information related to the semantic category of nouns. In the current version of USR guidelines, we are capturing information which is useful for generating languages under consideration. Those languages are Hindi, Tamil, English and Bangla. The categories are

1. Proper Nouns
2. Mass
3. Animacy
4. Definiteness
5. Inclusive/exclusive Pronominal entity

Proper Nouns: Following classes under this category will be annotated:

Proper Noun Class	**Abbreviation**
Person	per
Place (City, Continent)	place
Organization	org
Names of movies, medicine, cuisine, games, disease	misc

Mass: This tag 'maas' will be used for nouns which denote mass and uncountable objects. Ex. पानी, घास, चावल etc.

Animacy: Animate nouns are tagged with the label 'anim'.

Definiteness: If the noun denotes definiteness or specificity information, it is marked as 'def'.

When more than one semantic category needs to be represented for one entity, they can be separated with a semicolon (;).

Row1: मुख्य अतिथि ने विजेताओं को पुरस्कार दिये.

Row2: मुख्य अतिथि ने, विजेताओं को, पुरस्कार, दिये,.

Row4: def;anim, def;anim,,

Inclusive/Exclusive Pronominal Entity:

Languages like Tamil use separate lexical items for inclusiveness and exclusiveness in first person pronominal form. For example, the following Hindi sentence can have two interpretation:

हम मैच जीत गये

(i) The speaker is part of the team that has won the match

(ii) The speaker is not directly part of the team but he/she feels a relation with the team (for example, the team can be his/her school team)

Row 1: #राम ने बस अड्डे पर एक अच्छे लड़के के साथ ही बात की .

Row 2: #राम_ने,बस+अड्डा_पर,एक अच्छा लड़का_के_साथ,ही,बात+कर-या,.

Row 3: राम,बस+अड्डा_1,एक_1 अच्छा_1 लड़का_1,ही_1,बात+कर_1-या_1

Row 4: 1,2,3,4,5

Row 5: per,def,,,

Some clue for semantic category tagging:

If the noun is a definite noun and not preceded by a possessive pronoun or a demonstrative adjective or a determiner, write 'def'.

Row 6: GNP (Gender, Number, Person) Information

GNP information is mentioned for nouns, proper nouns and pronouns. It is written within [].

The value of Gender: m/f

where 'm' stands for masculine

'f' stands for feminine

Only inherent gender is marked and not grammatical gender. For example, बालक is inherently masculine. Therefore, we give the tag 'm'.

However, किताब remains unmarked although, in Hindi, this word is considered to be feminine.

The value of Number: sg/pl

where 'sg' stands for singular

'pl' stands for plural

The value of Person: u/m/a

where 'u' stands for उत्तम पुरुष (1st person)

'm' stands for मध्यम पुरुष (2nd person)

'a' stands for अन्य पुरुष (3rd person)

If any of the above fields are not necessary then mark the field as '-'

####Note: For 1st person gender information is not necessary. For 2nd person both gender and number fields can be left blank and marked as '-'. For nouns, gender information is not required.#####

Row 1: #राम ने बस अड्डे पर एक अच्छे लड़के के साथ ही बात की .

Row 2: #राम_ने,बस+अड्डा_पर,एक अच्छा लड़का_के_साथ, ही,बात+कर-या,.

Row 3: राम,बस+अड्डा_1,एक_1 अच्छा_1 लड़का_1, ही_1,बात+कर_1-या_1

Row 4: 1,2,3,4,5

Row 5: propn,def,,,

Row 6: [m sg a],[- sg a],[m sg a],,

Row 7: Intrachunk Dependency Relations

At Intrachunk level, the relations between a group head and its modifiers (dependents) are specified. Some relations are listed below:

Relation	**Abbreviation**

संख्या–विशेषण	ord
विशेषण	mod
परिमाणवाची–विशेषण	quant
क्रमवाची–विशेषण	card

The head and dependents are indexed internally within a group. For example, the intra–chunk index of the NP एक अच्छा लड़का_के_साथ is 3 in our example sentence. The index of each word in the group is given below:

एक	अच्छा	लड़का_के_साथ
3.1	3.2	3.3

Since लड़का is the head of this group, relations will be written down between लड़का and its modifiers एक and अच्छा shown below:

Row 1	राम ने बस अड्डे पर एक अच्छे लड़के के साथ ही बात की				
Row 2	राम_ने	बस+अड्डा_पर	एक अच्छा लड़का_के_साथ	ही	बात+कर–या
Row 3	राम	बस_1+अड्डा_1	एक_1 अच्छा_1 लड़का_1	ही_1	बात+कर_1–या_1
Row 4	1	2	3	4	5
Row 5	per	def	anim		
Row 6	[m sg a]	[– sg a]	[m sg a]		
Row 7		2.2:rt	3.3:ord;3.3:mod		

- The colon separates the head and dependency relation, the head being on the left of the colon. For example, 3.3 refers to the head लड़का_1 and 'ord' specifies the relation between एक_1 and लड़का_1.
- Semi–colon (;) separates the head:relation pair for each modifier

Row 8: Interchunk Relations

Four types of relations between the heads of two groups are presented here. They are:

- 1. **Karaka relation** between verbs and its dependent nouns
- 2. **Non-karaka relation** between verbs and its dependent nouns
- 3. Nominal head and its dependent in **sasThI relation**
- 4. Noun-Noun relation with **stative predicate**

1. Kaaraka relation between verbs and its dependent nouns

Relation	**Abbr.**
कर्ता	k1
कर्म	k2
करण	k3
संप्रदान	k4
अपादान	k5
स्थानाधिकरण	k7p
कालाधिकरण	k7t
अधिकरण	k7
मध्यस्थ–कर्ता	mk1
प्रयोजक–कर्ता	pk1
प्रयोज्य–कर्ता	jk1
उपादान कारण	k5mc (material cause)
Destination	k2p
अनुभव–कर्ता	k4a
frequency	k7tf (Ex. प्रतिदिन)
सह कर्ता	ras–k*

Row 1	राम ने बस अड्डे पर एक अच्छे लड़के के साथ ही बात की				
Row 2	राम_ने	बस+अड्डा_पर	एक अच्छा लड़का_के_साथ	ही	बात+कर–या
Row 3	राम	बस_1+अड्डा_1	एक_1 अच्छा_1 लड़का_1	ही_1	बात+कर_1–या_1

Row 4	1	2	3	4	5
Row 5	per	def	anim		
Row 6	[m sg a]	[− sg a]	[m sg a]		
Row 7		2.2:rt	3.3:ord;3.3:mod		
Row 8	5:k1	5:k7p	5:ras−k1		

2. Non-kaaraka relation between verbs and its dependent nouns

Relation	**Abbr**	
हेतु	rh	(cause relation)
तादर्थ्य	rt	(purpose relation)
सह	ras	(associate relation)
सादृश्य	ru	(upamAna relation)

Row 1	राम बस अड्डे पर बस के लिये प्रतीक्षा कर रहा है			
Row 2	राम	बस+अड्डा पर	बस के लिये	प्रतीक्षा कर रहा है
Row 3	राम	बस_1+अड्डा_1	बस_1	प्रतीक्षा+कर_1−0_ रहा_है_1
Row 4	1	2	3	4
Row 5	per	def		
Row 6	[m sg a]	[− sg a]	[− sg a]	
Row 7		2.2:rt		
Row 8	4:k1	4:k7p	4:rt	

4. Noun-Noun relation with stative predicate

When the verb is a stative predicate, the relation is specified not via the stative predicate but directly between the participants involved or between the participants and its attribute in case of copulative verbs.

How to distinguish between karaka and non−karaka relations between verbs and the dependent nouns

Non−karaka relations occur between the participants of the sentence and not between the verb and its participants.

i) AXAra−AXeya

Ex. राम दिल्ली में है (राम − AXeya; दिल्ली − AXAra)

ii) possessor−possessed

Ex. राम के पास किताब है (राम − possessor; किताब − possessed)

iii) anuBAvaka−anuBava

Ex. राम को बुखार है (राम − anuBAvaka; बुखार − anuBava)

iv) samAnAXi−samAnAXi

Ex. राम अच्छा है/राम एक छात्र है (राम, अच्छा both are samAnAXi in the 1st sentence; राम and छात्र are samAnAXi in the second.)

Note: अच्छा is not in samAnAXikaraNa relation with राम in राम एक अच्छा छात्र है as it is an attributive adjective. Only राम and छात्र are in samAnAXikaraNa relation with each other.

v) anuBAvaka−anuBAvya−anuBava

Ex. राम को श्याम के प्रति प्यार है (राम − anuBAvaka; श्याम − anuBAvya; प्यार − anuBava)

No verb id will be given for non−karaka relations.

Ex. Row 1: #rAma xillI meM hE.

Row 2: #rAma,xillI meM,hE,.

Row 3: rAma,xillI,state_existential−pres

Row 4: 1,2,3

...

Row 8:AXeya,AXAra,

...

Row 9: Discourse Relations

Relations such as emphasis, delimitation, co−referencing, negation are marked in this row. For discourse particles (dp), currently, we are not further specifying the meaning other than using the string itself to represent the semantics.

Discourse relation	**Abbreviation**
dp−hI	hI
dp−BI	BI
dp−wo	wo
negation	neg
co−reference	co−ref
dp−waka	waka

Row 1	राम ने बस अड्डे पर एक अच्छे लड़के के साथ ही बात की				
Row 2	राम_ने	बस+अड्डा_पर	एक अच्छा लड़का_के_साथ	ही	बात+कर−या
Row 3	राम	बस_1+अड्डा_1	एक_1 अच्छा_1 लड़का_1	ही_1	बात+कर_1−या_1
Row 4	1	2	3	4	5
Row 5	per	def	anim		
Row 6	[m sg a]	[− sg a]	[m sg a]		
Row 7		2.2:rt	3.3:ord;3.3:mod		
Row 8	5:k1	5:k7p	5:ras−k1		
Row 9				3:hI	

इन तीन लडके को जाना है

इन तीन लडके को ही जाना है

इन तीन ही लडके को जाना है

इन ही तीन लडके को जाना है

Negation:

राम ने बस अड्डे पर एक अच्छे लड़के के साथ बात नहीं की.

Row 1	राम ने बस अड्डे पर एक अच्छे लड़के के साथ बात नहीं की				
Row 2	राम_ने	बस+अड्डा_पर	एक अच्छा लड़का_के_साथ	नहीं	बात+कर-या
Row 3	राम	बस_1+अड्डा_1	एक_1 अच्छा_1 लड़का_1	नहीं_1	बात+कर_1-या_1
Row 4	1	2	3	4	5
Row 5	per	def	anim		
Row 6	[m sg a]	[- sg a]	[m sg a]		
Row 7		2.2:rt	3.3:ord;3.3:mod		
Row 8	5:k1	5:k7p	5:ras-k1		
Row 9				4:neg	

Row 10: Sentence Types

The user has to mention the sentence type in the 9th row.

Examples of sentence types:

राम ने बस अड्डे पर एक अच्छे ही लड़के के साथ बात की - declarative (positive)

राम आज स्कूल नहीं गया - declarative (negative)

क्या तुमने एक नयी गाड़ी खरीदी है? - question

तुम घर जाओ - imperative

Following are some more examples:

राम ने तीन किलो चक्की आटा दस तारीख को खरीदा

Row 1	राम ने	बस अड्डे पर	एक अच्छे लड़के के साथ	ही	बात की
Row 2	राम_ने	बस+अड्डा_पर	एक अच्छा लड़का_के_साथ	ही	बात+कर-या
Row 3	राम	बस_1+अड्डा_1	एक_1 अच्छा_1 लड़का_1	ही_1	बात+कर_1-या_1
Row 4	1	2	3	4	5
Row 5	per	def	anim		
Row 6	[m sg a]	[- sg a]	[m sg a]		
Row 7		2.2:rt	3.3:ord;3.3:mod		
Row 8	5:k1	5:k7p	5:ras-k1		
Row 9				3:hI	
Row 10	assertive				

Examples of different types of sentences and USRs

Stative verbs:

i) Copulative

#rAma dAktara hE.
#rAma_0,dAktara_0,hE,.
rAma,dAktara_1,state_copula-pres
1,2,3
propn,,
[m sg a],[- sg a],
,,
samAnAXi,samAnAXi,
,,
assertive

ii) predicative adjective

#rAma acCA hE.
#rAma_0,acCA_0,hE,.
rAma,acCA_1,state_copula-pres
1,2,3
propn,guNavAcI,
[- sg a],,

„
samAnAXi,samAnAXi,
„
assertive

iii) existential

#rAma xillI meM hE.
#rAma_0,xillI_meM,hE,.
rAma,xillI,state_existential-pres
1,2,3
propn,propn,
[m sg a],[- sg a],
„
AXeya,AXAra,
„
assertive

iv) experiencer (anuBAvaka-anuBava relation)

#rAma ko buKAra hE.
#rAma_ko,buKAra_0,hE,.
rAma,buKAra_1,state_anuBUwi-pres
1,2,3
propn,mass,
[m sg a],[- sg a],
„
anuBAvaka,anuBava,
„
assertive

v) experiencer (anuBAvaka-anuBAvya-anuBava relation)

#rAma ko SyAma se pyAra hE.
#rAma_ko,SyAma_se,pyAra_0,hE,.
rAma,SyAma,pyAra_1,state_anuBUwi-pres
1,2,3,4
propn,propn,mass,
[m sg a],[m sg a],[- sg a],
„„
anuBAvaka,anuBAvya,anuBava,
„„
assertive

vi) possession

#rAma ke pAsa kiwAba hE.

#rAma_ke_pAsa,kiwAba_0,hE.
rAma,kiwAba_1,state_possession-pres
1,2,3
propn,def,
[m sg a],[- sg a],
„
possessor,possessed,
„
assertive