

-----USER CSV GUIDELINES-----

#####

The user csv file will contain 8 rows. The rows are

- 1: Grouping
- 2: Concept from Concept dictionary
- 3: Index for Concepts
- 4: Lexical Category (only for proper nouns, pronouns) AND Ontological Information about nouns
- 5: GNP (Gender, Number, Person) Information
- 6: Intrachunk Dependency Relations
- 7: Interchunk Dependency Relations
- 8: Discourse Relations

The entries of each row is separated with the comma (',').
For a sentence "राम ने बस-अड़डे पर एक अच्छे ही लड़के के साथ बात की", we will build the information of each row as we describe them below

Row 1 : Grouping

For an introduction of Grouping see Appendix 1.
For an introduction of TAM and the list of TAM see Appendix 2

Nouns - The relation between a noun and a verb or another noun are either unmarked or marked by a post-position.
When unmarked, zero is postulated, otherwise postpositions are grouped with the noun by an underscore.
राम_ने, फल_0, घर_के_पास

Compound Nouns - The constituent nouns are joined with a '+'.
बस्+अड़डा (bus-stop)

Verbs - Verbs can be of tinangta (finite) form or kridanta (non-finite, verbal noun, participial) form.
re tinangta:

करेगा, कर रहा है, किया था

Verb can be split into root and TAM. The TAM can be
(i) Attached with the verb as in करेगा ; or
(ii) Written separately when the root can be either
(a) In a bare form as in कर रहा है ; or
(b) Not bare as in किया था

The TAM is separated from the root by '-'. 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, कर-0_रहा_है कर-या_था

है/था, थी are irregular verbs in the language. There are represented as it is.

krdanta:

खा कर खाया हुआ खाते खाते
खा_कर खा-या_हुआ खा-ते खा-ते

Negation - Negation indicators should be written after the verb.

e.g. रहा_है, नहीं

Compound Verbs -

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

For स्नान किया - स्नान+कर-या

Pronouns -

तुम्हारे, मैं, मुझे, मेरा
तुम_का, मैं_0, मैं_को, मैं_का

Adjectives, Quantifier and संख्या/number are grouped with nouns and the symbol used is ~.

e.g. एक~अच्छा*ही~लड़का_के_साथ

Adjectives - अच्छा~लड़का, छोटा~लड़का

Quantifier - बहुत~काम

Adverbs - धीरे

Particles - तक, ही, भी

Particles will be joined by '*'. e.g घर*तक, राम*ही, तुम*भी

Other words such as 'साहब', 'जी' which denote respect will also be joined by '*'.

saMKyA (number) - A number can be followed by a measuring unit or by noun or (viSeRaNa+noun). When the following word is a measuring unit then join by "#".

e.g. चार#किलोमीटर, 4#किलोमीटर

When the following word is noun or (viSeRaNa+noun), then join the number by '@'.

e.g. 4@लड़का, चार@लड़का, चार@छोटा~लड़का

परिमाणवाची (Measurement words) - किलोमीटर, मीटर, ग्राम

If the word is a measuring unit (see the list), then join the unit with the following noun or (viSeRaNa+noun) by '@'.

e.g. चार#किलोमीटर@लम्बा~रस्ता, एक#दर्जन@केला.

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

क्रमवाची words are joined with '~' as adjectives. e.g. पहला~घर

Comparatives - तर, तम, सबसे

Comparatives are joined with '^'. e.g. सबसे^अच्छा, सुन्दर^तम, सुन्दर^तर

Connectives - कि, अगर, तो

Constructions: -

Noun conjunctions - और, या, आदि

If two or more nouns are joined by the conjunctions, then we have to make the construction format as shown below:

sent: राम, श्याम और सीता को विद्यालय जाना चाहिए

construction: {[राम][श्याम][सीता]_को]और} विद्यालय_0 जा-ना_चाहिए

sent: राम, श्याम और सीता विद्यालय जा रहे हैं

construction: {[राम][श्याम][सीता]और} विद्यालय_0 जा-0_रहा_है

All the nouns/pronouns/prop nouns which are joined with the conjunction(s), will be written in separate "[" and the vibhakti (if any) will be added after all the nouns/pronouns/prop nouns. At the end of the construction, और will be added within "{}".

End of the sentence - The punctuation occurs at the last column of the row 1.
राम_ने, बस+अड़ड़ा_पर, एक~अच्छा*ही~लड़का_के_साथ, बात+कर-या, .

Row 1: राम_ने, बस+अड़ड़ा_पर, एक~अच्छा*ही~लड़का_के_साथ, बात+कर-या, .

Row 2: Concept from Concept dictionary

For an introduction of Concept Dictionary see Appendix 3.

The prakrti (base form) of row 1 is substituted by a concept. In other words, the word is disambiguated. For the correct concept id, check the concept dictionary.

Note: a. For Proper nouns the sense id will always be '1'.
b. No id will be used for pronouns.

Row 2:

राम_ने, बस+अड़ड़ा_पर, एक~अच्छा*ही~लड़का_के_साथ, बात+कर-या, .
राम_1, बस+अड़ड़ा_1, एक_1~अच्छा_1~लड़का_1, बात+कर_1

Row 3: Index for Concepts

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

Row 3:

राम_ने, बस+अड़ड़ा_पर, एक~अच्छा*ही~लड़का_के_साथ, बात+कर-या, .
राम_1, बस+अड़ड़ा_1, एक_1~अच्छा_1~लड़का_1, बात+कर_1
1, 2, 3, 4

Row 4: Lexical Category

Only the categories such as proper noun, pronoun will be provided in the 4th row.

The abbreviations of categories are given below

Category	Abbr.	
noun	n	
verb	v	
pronoun	pron	#Ex: वह, मैं, तुम, आप
proper_noun	propn	#Ex: राम, सीता
adverb	adv	#Ex: धीरे
adjective	adj	#Ex: अच्छा, छोटा
time adv	कालवाची	#Ex: कब
place adv	स्थानवाची	#Ex: कहाँ
wh_pronoun	wh-pron	#Ex: क्या, कौन
conjunction	c	#Ex: और, या
quantifiers	det	#Ex: बहुत
mass noun	massn	#Ex: चावल, चीनी, पानी
cardinal nos.	संख्या	#Ex: दो, 2
compound noun	comp_n	#Ex: बस-अड़ड़ा
relative pron	rel_pron	#Ex: जो
quasi-compound	quasi_comp	#Ex: एक~अच्छा~लड़का

Row 4:

राम_ने, बस+अड़ड़ा_पर, एक~अच्छा*ही~लड़का_के_साथ, बात+कर-या, .
राम_1, बस+अड़ड़ा_1, एक_1~अच्छा_1~लड़का_1, बात+कर_1

1,2,3,4
propn,def,,

Ontological Information about nouns

For an introduction of Ontological Information see Appendix 4.

If the noun is a definite noun and not preceded by a possessive pronoun or a determiner, write 'def'. Otherwise leave it blank as it is. If the noun is a mass noun then write 'mass'.

Row 4:

राम_ने, बस+अड़ड़ा_पर, एक~अच्छा*ही~लड़का_के_साथ, बात+कर-या, .
राम_1, बस+अड़ड़ा_1, एक_1~अच्छा_1~लड़का_1, बात+कर_1
1,2,3,4
propn,def,,

Row 5: 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/n
 where 'm' stands for masculine
 'f' stands for 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 (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 5:

राम_ने, बस+अड़ड़ा_पर, एक~अच्छा*ही~लड़का_के_साथ, बात+कर-या, .
राम_1, बस+अड़ड़ा_1, एक_1~अच्छा_1~लड़का_1, बात+कर_1
1,2,3,4
propn,def,,
[m sg a],[- sg a],[m sg a],

Row 6: Intrachunk Dependency Relations

For an introduction of Dependency Relation see Appendix 5 and for an introduction of Intrachunk see Appendix 5.1.

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

Relation	Abbr.
संख्या-विशेषण	संख्या-विशेषण
विशेषण	विशेषण
परिमाणवाची-विशेषण	परिमाणवाची
क्रमवाची-विशेषण	क्रमवाची

The head and dependents are indexed internally. For example, the intra-chunk index of the NP एक~अच्छा_ही~लड़का_के_साथ which as index 3 (see below in the example) is:

एक~अच्छा*ही~लड़का_के_साथ
3.1~3.2~3.3

The relation is written for the child with the intra-chunk index id of its head specified:

Row 6: राम_ने, बस+अड़डा_पर, एक~अच्छा*ही~लड़का_के_साथ, बात+कर-या, .
राम_1, बस+अड़डा_1, एक_1~अच्छा_1~लड़का_1, बात+कर_1
1, 2, 3, 4
propn, def, ,
[m sg a], [- sg a], [m sg a],
, , 3.3:संख्या-विशेषण~3.3:विशेषण,

Row 7: Interchunk Relations

a. kaarak relation between a verb and its participants:

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
सह-कर्ता	ras-k1
destiantion	k2p
अनुभव-कर्ता	k4a
कर्ता-समानाधिकरण	k1s
कर्म-समानाधिकरण	k2s

c. Relation between two verbs within a sentence:

Relation	Abbr.
if-then relation	condition
पूर्व_कालिक_क्रिया	पूर्व_कालिक
सम_कालिक_क्रिया	सम_कालिक

Row 7:

राम_ने, बस+अड़डा_पर, एक~अच्छा*ही~लड़का_के_साथ, बात+कर-या, .
राम_1, बस+अड़डा_1, एक_1~अच्छा_1~लड़का_1, बात+कर_1
1, 2, 3, 4
propn, def, ,
[m sg a], [- sg a], [m sg a],
, , 3.3:संख्या-विशेषण~3.3:विशेषण,
4:k1, 4:स्थानवाची, ras-k1,

Row 8: Discourse Relations

Relations such as emphasis, delimitation, co-refencing are marked in this row.

Relation	Abbr.
emphasis	emph
boundary/	सीमा
delimitation	
Co-ref	यत्-तत्

Row 8:

राम_ने, बस+अड़डा_पर, एक~अच्छा*ही~लड़का_के_साथ, बात+कर-या, .
राम_1, बस+अड़डा_1, एक_1~अच्छा_1~लड़का_1, बात+कर_1
1, 2, 3, 4
propn, def, ,
[m sg a], [- sg a], [m sg a],
, , 3.3:संख्या-विशेषण~3.3:विशेषण,
4:k1, 4:स्थानवाची, 4:ras-k1,
, , 3.2:emp,