

Course Project Report

Instructor: Dr. Parameshwari

Name: Priyank Modi, Roll Number: 20171055

Instructions: Read all the instructions below carefully before you start working on the report, and before you make a submission.

- Please typeset your submissions in \LaTeX . Use the template provided for your answers. Please include your name and Roll number with submission.
- This report is due on 27 April at 11:55 PM.
- No extensions will be given for the submission under any circumstances.
- Submissions via any other method other than Moodle submission will be deemed invalid.
- Plagiarism of any sort will not be tolerated. Strict action will be taken for those caught in plagiarism.

1 Tags

Discuss the different kind of tags in **both** UD and AnnCorra occurring in your chosen Indian language.

1.1 UD

The head has been underlined while the dependent has been written in **Bold**. The relations mentioned are in descending order of occurrence(based on the results from section 3, verified with those of the UD treebank for Hindi)

Tag	Examples	Linguistic Cues
nsubj	<ul style="list-style-type: none"> • buddh <u>ne</u> apne antim vishra:m ke lie kusha:vati: ko chuna: • priyanka: zi:nat ama:n wa:li: bhu:mika: nibha:e(n)gi:(n) • iss shEhEr ke 5 shata:badi: pura:ne hone ka prama:n milta: hE. 	<ul style="list-style-type: none"> • Relations : VERB-NOUN, VERB-PROPN, VERB-PRON • Multiple Cue 2
nsubj:pass	<ul style="list-style-type: none"> • Ra:m ti:m kapta:n bana:ya: gaya: hE • kal sha:m cha:r jawa:n ma:re gaye 	<ul style="list-style-type: none"> • Relations : VERB-NOUN, VERB-PROPN, VERB-PRON • Multiple Cue 2
obj	<ul style="list-style-type: none"> • isi mEda:n me(n) mEch khela: gaya: tha: • koi: bhi: ise pakad nahi(n) sakta: • mai(n) uske sa:th ga:na ga:na cha:hta: hu(n) 	<ul style="list-style-type: none"> • Relations : VERB-NOUN, VERB-VERB, VERB-PRON • Multiple Cue 2
iobj	<ul style="list-style-type: none"> • mujhe bohut ka:m mila: hE • usne Kejriwa:l ko cha:(n)ta: ma:ra: • use ja:ne do 	<ul style="list-style-type: none"> • Relations : VERB-NOUN, VERB-PROPN, VERB-PRON • Multiple Cue 2

obl	<ul style="list-style-type: none"> • iss shEhEr ke 5 shata:badi: pura:ne hone ka prama:n milta: hE • samjho:ta: amri:ki: sarka:r ke sa:th kiya: gaya: hE • yaha:(n) par mEch khela: gaya: tha: 	<ul style="list-style-type: none"> • Relations : VERB-NOUN, VERB-PROPN, VERB-PRON • Multiple Cue 2
vocative	<ul style="list-style-type: none"> • aaiye chale(n) Dilli • ve bole - makka:r kahi:(n) ke, nikal ja: desk se 	<ul style="list-style-type: none"> • Relations : VERB-NOUN, VERB-VERB, NOUN-NOUN • Multiple Cue 2
advcl	<ul style="list-style-type: none"> • Eksi:dE(n)t tab hua:, jab ra:t hone wa:li: thi: • ma(n)dir ja:ne ke lie pul ka nirman kia: gaya: hE 	<ul style="list-style-type: none"> • Relations : VERB-VERB • Multiple Cue 2
aux	<ul style="list-style-type: none"> • iss shEhEr ke 5 shata:badi: pura:ne hone ka prama:n milta: hE. 	<ul style="list-style-type: none"> • Relations : VERB-AUX • Multiple Cue 2
aux:pass	<ul style="list-style-type: none"> • vaha:(n) par ma(n)dir ka udhgha:tan kia: gaya: tha:. 	<ul style="list-style-type: none"> • Relations : VERB-AUX • Multiple Cue 2
cop	<ul style="list-style-type: none"> • Ram budhima:n hE. • ye pa:rti: ma:rti(n) lu:thar ki(n)g se prerit hE. 	<ul style="list-style-type: none"> • Relations : NOUN-AUX, ADJ-AUX • Multiple Cue 2
mark	<ul style="list-style-type: none"> • iss shEhEr ke 5 shata:badi: pura:ne hone ka prama:n milta: hE. • ba:t yah hE ki use saza: nahi(n) milni: cha:hie 	<ul style="list-style-type: none"> • Relations : VERB-ADB, VERB-SCONJ • Multiple Cue 2
nmod	<ul style="list-style-type: none"> • ghar ki: <u>khiDkiya:(n)</u> bahut baDi: hE(n). • Mi:na: ke <u>ba:l</u> bahut sundar hE(n). • unki: <u>U:(n)cha:i:</u> kari:b 15 m:tar hE. 	<ul style="list-style-type: none"> • Relations : NOUN-NOUN, NOUN-PROPN, NOUN-PRON, PRON-NOUN, PRON-PRON • Multiple Cue 2
nummod	<ul style="list-style-type: none"> • sher ke do <u>bacche</u> hE(n). 	<ul style="list-style-type: none"> • Relations : NOUN-NUM • Multiple Cue 2

acl	<ul style="list-style-type: none"> • Dilli: se a:ya: vah <u>laDka:</u> bahut hoshi:ya:r hE. • ab yah ja:nana: ba:ki hE ki voh yaha:(n) kyu:(n) a:ya: <u>hE</u>. 	<ul style="list-style-type: none"> • Relations : NOUN-VERB, PRON-VERB • Multiple Cue 2
acl:recl	<ul style="list-style-type: none"> • mE(n) us din ki: ba:t kar raha: hu:(n) jab vah pEhli: ba:r <u>dikha:</u> tha: • jab kashmi:ri: pa(n)Dito(n) par hamla: hua: <u>to</u> tum kaha:n the? 	<ul style="list-style-type: none"> • Relations : NOUN-VERB, PRON-VERB • Multiple Cue 2
amod	<ul style="list-style-type: none"> • mE(n)ne ek sundar <u>laDki:</u> dekhi: 	<ul style="list-style-type: none"> • Relations : NOUN-ADJ • Multiple Cue 2
det	<ul style="list-style-type: none"> • pra:chi:n ka:l me(n) is <u>stha:n</u> ka: bahut mahatva hE. 	<ul style="list-style-type: none"> • Relations : NOUN-DET • Multiple Cue 2
case	<ul style="list-style-type: none"> • main unhe <u>bhadgwa:n</u> ka banda: ma:nta: hu:(n). • bhagva:n <u>Ra:m</u> ke putra kush bahut balsha:li: the. 	<ul style="list-style-type: none"> • Relations : NOUN-ADP, PROP-ADP • Multiple Cue 2
advmod	<ul style="list-style-type: none"> • Esa: karne se sukh ki: pra:pati: nahi:(n) <u>hoti:.</u> • sabhi: utsukta: se inteza:r <u>kar</u> rahe hE(n). 	<ul style="list-style-type: none"> • Relations : VERB-PART, VERB-ADV • Multiple Cue 2
conj	<ul style="list-style-type: none"> • aaj <u>Ra:m</u> aur Shya:m kyu:(n) nahi:(n) a:ye?. • ba:za:r se <u>a:lu:</u> Or pya:z le a:na:. 	<ul style="list-style-type: none"> • Relations : PROP-PROP, NOUN-NOUN, VERB-VERB • Multiple Cue 2
cc	<ul style="list-style-type: none"> • yah andar se narm hE Or ba:har se rang <u>chhoDta:</u> hE. • Ra:m Or <u>Shya:m</u>, dono ko bula:o. 	<ul style="list-style-type: none"> • Relations : VERB-CCONJ, NOUN-CCONJ, PROP-CCONJ • Multiple Cue 2

compound	<ul style="list-style-type: none"> Ga:(n)dhi: ji: ka: janm 1869 <u>i:svi:</u> me(n) hua:. a:i:e <u>kare(n)</u> sEr. ma(n)dir <u>parisar</u> me(n) khaDe a:dmi: par sabka: dhya:n tha:. 	<ul style="list-style-type: none"> Relations : PROPN-PROP, NOUN-NOUN, VERB-ADJ Multiple Cue 2
root	<ul style="list-style-type: none"> a:i:e kare(n) sEr. 	<ul style="list-style-type: none"> Relations : -VERB Multiple Cue 2
dep	<ul style="list-style-type: none"> iss <u>shEhEr</u> ke bhi: 5 shata:badi: pura:ne hone ka prama:n milta: hE. <u>Ra:m</u> ko bhi: iske ba:re me(n) nahi:(n) pata:. 	<ul style="list-style-type: none"> Relations : NOUN-PART, PROP, PART, PRON-PART Multiple Cue 2
dislocated	<ul style="list-style-type: none"> kal hi: Ra:m <u>Bana:ras</u> (Va:ra:nasi:) se a:ya 	<ul style="list-style-type: none"> Relations : PROP-ADP, NOUN-ADP Multiple Cue 2

Table 1: Discussion on UD tags

1.2 AnnCorra

The head has been underlined while the dependent has been written in **Bold**. The relations mentioned are in descending order of occurrence(based on the results from section 3, verified with those of the UD treebank for Hindi)

Tag	Examples	Linguistic Cues
k1	<ul style="list-style-type: none"> Ra:m vaha:n khada: hE. Ra:m ne kha:na: kha:ya: iss <u>shEhEr</u> ke 5 shata:badi: pura:ne hone ka prama:n milta: hE. 	<ul style="list-style-type: none"> Type : NOUN-VERB Postpositions : NULL, ne, ko
nsubj:pass	<ul style="list-style-type: none"> Ra:m ti:m kapta:n bana:ya: gaya: hE kal sha:m cha:r jawa:n ma:re gaye 	<ul style="list-style-type: none"> Relations : VERB-NOUN, VERB-PROP, VERB-PRON Multiple Cue 2
obj	<ul style="list-style-type: none"> isi mEda:n me(n) mEch khela: gaya: tha: koi: bhi: ise pakad nahi(n) sakta: mai(n) uske sa:th ga:na ga:na cha:hta: hu(n) 	<ul style="list-style-type: none"> Relations : VERB-NOUN, VERB-VERB, VERB-PRON Multiple Cue 2

iobj	<ul style="list-style-type: none"> • mujhe bohot ka:m mila: hE • usne Kejriwa:l ko cha:(n)ta: ma:ra: • use ja:ne do 	<ul style="list-style-type: none"> • Relations : VERB-NOUN, VERB-PROPN, VERB-PRON • Multiple Cue 2
obl	<ul style="list-style-type: none"> • iss shEhEr ke 5 shata:badi: pura:ne hone ka prama:n milta: hE • samjho:ta: amri:ki: sarka:r ke sa:th kiya: gaya: hE • yaha:(n) par mEch khela: gaya: tha: 	<ul style="list-style-type: none"> • Relations : VERB-NOUN, VERB-PROPN, VERB-PRON • Multiple Cue 2
vocative	<ul style="list-style-type: none"> • aaiye chale(n) Dilli • ve bole - makka:r kahi:(n) ke, nikal ja: desk se 	<ul style="list-style-type: none"> • Relations : VERB-NOUN, VERB-VERB, NOUN-NOUN • Multiple Cue 2
advcl	<ul style="list-style-type: none"> • Eksi:dE(n)t tab hua:, jab ra:t hone wa:li: thi: • ma(n)dir ja:ne ke lie pul ka nirman kia: gaya: hE 	<ul style="list-style-type: none"> • Relations : VERB-VERB • Multiple Cue 2
aux	<ul style="list-style-type: none"> • iss shEhEr ke 5 shata:badi: pura:ne hone ka prama:n milta: hE. 	<ul style="list-style-type: none"> • Relations : VERB-AUX • Multiple Cue 2
aux:pass	<ul style="list-style-type: none"> • vaha:(n) par ma(n)dir ka udhgha:tan kia: gaya: tha:. 	<ul style="list-style-type: none"> • Relations : VERB-AUX • Multiple Cue 2
cop	<ul style="list-style-type: none"> • Ram budhima:n hE. • ye pa:rti: ma:rti(n) lu:thar ki(n)g se prerit hE. 	<ul style="list-style-type: none"> • Relations : NOUN-AUX, ADJ-AUX • Multiple Cue 2
mark	<ul style="list-style-type: none"> • iss shEhEr ke 5 shata:badi: pura:ne hone ka prama:n milta: hE. • ba:t yah hE ki use saza: nahi(n) milni: cha:hie 	<ul style="list-style-type: none"> • Relations : VERB-ADB, VERB-SCONJ • Multiple Cue 2
nmod	<ul style="list-style-type: none"> • ghar ki: <u>khiDkiya:(n)</u> bahut baDi: hE(n). • Mi:na: ke <u>ba:l</u> bahut sundar hE(n). • unki: <u>U:(n)cha:i:</u> kari:b 15 m:tar hE. 	<ul style="list-style-type: none"> • Relations : NOUN-NOUN, NOUN-PROPN, NOUN-PRON, PRON-NOUN, PRON-PRON • Multiple Cue 2

nummod	<ul style="list-style-type: none"> sher ke do <u>bacche</u> hE(n). 	<ul style="list-style-type: none"> Relations : NOUN-NUM Multiple Cue 2
acl	<ul style="list-style-type: none"> Dilli: se a:ya: vah <u>laDka:</u> bahut hoshi:ya:r hE. ab yah ja:nana: ba:ki hE ki voh yaha:(n) kyu:(n) a:ya: <u>hE</u>. 	<ul style="list-style-type: none"> Relations : NOUN-VERB, PRON-VERB Multiple Cue 2
acl:recl	<ul style="list-style-type: none"> mE(n) us din ki: ba:t kar raha: hu:(n) jab vah pEhli: ba:r <u>dikha:</u> tha: jab kashmi:ri: pa(n)Dito(n) par hamla: hua: <u>to</u> tum kaha:n the? 	<ul style="list-style-type: none"> Relations : NOUN-VERB, PRON-VERB Multiple Cue 2
amod	<ul style="list-style-type: none"> mE(n)ne ek sundar <u>laDki:</u> dekhi: 	<ul style="list-style-type: none"> Relations : NOUN-ADJ Multiple Cue 2
det	<ul style="list-style-type: none"> pra:chi:n ka:l me(n) is <u>stha:n</u> ka: bahut mahatva hE. 	<ul style="list-style-type: none"> Relations : NOUN-DET Multiple Cue 2
case	<ul style="list-style-type: none"> main unhe <u>bhadgwa:n</u> ka banda: ma:nta: hu:(n). bhagva:n <u>Ra:m</u> ke putra kush bahut balsha:li: the. 	<ul style="list-style-type: none"> Relations : NOUN-ADP, PROP-ADP Multiple Cue 2
advmod	<ul style="list-style-type: none"> Esa: karne se sukh ki: pra:pati: nahi:(n) <u>hoti:.</u> sabhi: utsukta: se inteza:r <u>kar</u> rahe hE(n). 	<ul style="list-style-type: none"> Relations : VERB-PART, VERB-ADV Multiple Cue 2
conj	<ul style="list-style-type: none"> aaj <u>Ra:m</u> aur Shya:m kyu:(n) nahi:(n) a:ye?. ba:za:r se <u>a:lu:</u> Or pya:z le a:na:. 	<ul style="list-style-type: none"> Relations : PROP-PROP, NOUN-NOUN, VERB-VERB Multiple Cue 2
cc	<ul style="list-style-type: none"> yah andar se narm hE Or ba:har se rang <u>chhoDta:</u> hE. Ra:m Or <u>Shya:m</u>, dono ko bula:o. 	<ul style="list-style-type: none"> Relations : VERB-CCONJ, NOUN-CCONJ, PROP-CCONJ Multiple Cue 2

compound	<ul style="list-style-type: none"> Ga:(n)dhi: ji: ka: janm 1869 i:svi: me(n) hua:. a:i:e kare(n) sEr. ma(n)dir <u>parisar</u> me(n) khaDe a:dmi: par sabka: dhya:n tha:. 	<ul style="list-style-type: none"> Relations : PROP-N-PROP-N, VERB-NOUN, NOUN-NOUN, VERB-ADJ Multiple Cue 2
root	<ul style="list-style-type: none"> a:i:e kare(n) sEr. 	<ul style="list-style-type: none"> Relations : -VERB Multiple Cue 2
dep	<ul style="list-style-type: none"> iss shEhEr ke bhi: 5 shata:badi: pura:ne hone ka prama:n milta: hE. <u>Ra:m</u> ko bhi: iske ba:re me(n) nahi:(n) pata:. 	<ul style="list-style-type: none"> Relations : NOUN-PART, PROP-N-PART, PRON-PART Multiple Cue 2
dislocated	<ul style="list-style-type: none"> kal hi: Ra:m <u>Bana:ras</u> (Va:ra:nasi:) se a:ya 	<ul style="list-style-type: none"> Relations : PROP-N-ADP, NOUN-ADP Multiple Cue 2

Table 2: Discussion on UD tags

1.3 AnnCorra

Do the same for AnnCorra.

2 Linguistic Challenges with Annotation

In this section, first an overview of the phenomenon along with some examples is given before proceeding to what challenges occur because of these during annotation.

(a) Differential Object marking

DOM is largely grammaticalized in Modern Standard Hindi, where identified objects are both case marked and can trigger a change in verb agreement: in ergative constructions, 1 as well as in passive constructions, the verb agrees with an unmarked patient, but not with a marked patient. From my observations, DOM occurs only with formally transitive verbs and this makes sense if we look at the kind of markers used for transitive verbs. A quick examples to show the difference between the default agreement and agreement with unmarked objects is given below:

mEne **is film ko** dekha:.
maine **yah film** dekhi.

Getting into a little more detail, there are 3 main arguments which lead to a change in the marking of direct objects viz Animacy, definiteness and specificity(which can be integrated into the previous). The following are the observations in a bit more detail :

- Animacy with an indefinite human object is obligatorily marked, and so are proper nouns referring to human objects.

kisi: **ko** bula:o.

- The opposite happens with those referring to inanimate objects.

mE ne Dilip ko (*Dilip) dekha:/ mEne Kalkatta: dekha:.

- Pronominalized inanimate objects are more often marked than the corresponding nouns

koi: pita: bhi: ise barda:sht nahi:(n) kar sakta:.

While I've made an attempt to list the arguments and the changes they bring to object marking, in a lot of cases it's hard to capture the disappearance of the object case marker which makes it very difficult to build systems that search for the object. For instance in

Mi:na: ne ba:za:r mein ek bandar ko nachte hue dekha.
Mi:na: ne ba:za:r mein ek bandar nachte hue dekha.

In the second example, Mina is doing the act of dancing. So, the non-finite adjunct 'dancing' can only be controlled by the subject of the matrix clause Mina, not by the unmarked object, whereas the same, when marked, controls the adjunct. However, unmarked objects are commonly used with an adjunct that they control, although they are in this case typically inanimate. It is very hard to pick the object because of the absence of the case marker 'ko' here. It was also hard to pick unmarked objects in participial constructions especially when the object came in sentence-initial position.

gaari waapas aati hui dekhkar maine socha.
kaale gulab jamun dekhkar uske muh se laar tapakne lagi.

(b) Non-Nominative Subjects

Non-Nominative Subjects is a phenomenon, which as the name suggests, where certain verbs specify that their subjects are to be in a case other than nominative. Finite verb agreement in Hindi is found only with nominative DPs. For non nominative DPs, it's often tough for the parser to identify the subject. For instance :

Ram ko apni pustak mili

Here subject Ram exists in the Dative case and hence is hard to identify for a UD tagger.

These non-nominative subjects in Hindi pass most of the test for subjecthood as well like subject-oriented anaphora binding, PRO control, reduced relative clause, conjunction reduction,[3] subject-to-subject raising, and subject-to-object raising; but still exist in the non-nominative case. This poses problems while using UD tags which are more affected by positioning and less by case markers because of their more syntactical definition for tags. However when it comes to AnnCorra, since it relies much more on semantics and thus case markers, it's easy to capture the correct tags.

(c) Complex Predicates

Complex Predicates occur in the form of nominal+verb combinations (called conjunct verbs) and verb+verb combinations (called compound verbs). It becomes really hard to tag such instances. As per my observations, we can broadly divide these problems into the following two:

- Given a conjunct verb i.e. a Noun+Verb combination, is the noun incorporated into the verb complex or is it an overt argument of the verb?

Ram ne kal shaadi ki.

Here, it's hard to decide for a tagger (and even for human annotators) whether shaadi is to be an argument of the root verb ki or does it need to be marked together with ki as a local word group. This is handled very differently in UD and AnnCorra and personally I found it hard to find the correct tags in such cases especially while marking AnnCorra because of the existence of such a huge number of semantic relations.

- Given a Compound Verb i.e. a Verb+Verb combination, is the second verb an aspectual/modal or is it the polar (intensifier) in a polar-vector combination?

Ram ghar chala gaya.

While the problem of guessing if chala is a modal or an intensifier or neither exists, there are other problems with complex verbs in Hindi. V+V constructions are also difficult to deal with and describe, since there are many serial verb sequences to be found in the language. Consider,

Ram kitaab parh raha hai.
Ram ne kitaab parh daali.

rahaa in the first sentence marks the progressive aspect, whereas daalii in second one indicates the attitude (somehow finished). The V+V sequence in the former should not be stored in the lexicon, whereas the one in latter should probably be.

(d) Non-finite clauses: Conditional, Concessive, Relative, participial clauses

A Non-Finite verb does not have subject, tense or number and it does not change its form according to the number and person of the subject. Conditional, Concessive, Relative clauses do not seem related to the concept of finiteness of verbs. For instance, for the concessive clause :

halanki koi nahi uski baat sun raha tha, vah fer bhi baat karta raha
 “Although no one was listening to him, he kept talking anyway”

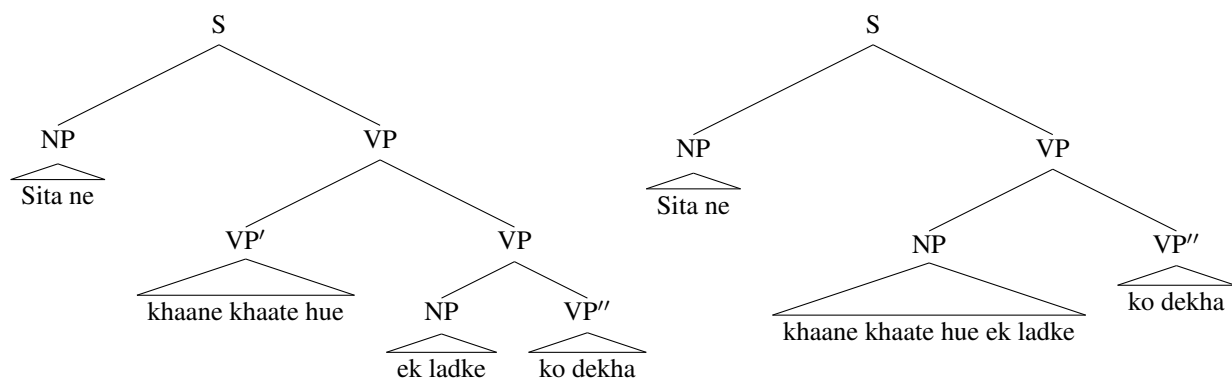
there seems to be no ambiguity. More importantly the verb is finite. Since concessive constructions are pretty restrictive in the sense that the word ‘although’ or its forms should exist, i can’t think of any examples where concessive clauses are non finite and lead to ambiguity. Same goes for conditional and Relative clauses. The only part that makes sense under this heading is Participial Clauses, atleast for Hindi. Could you please explain what we are expected to write for this part?

(e) Ambiguity (Coordination, Attachment)

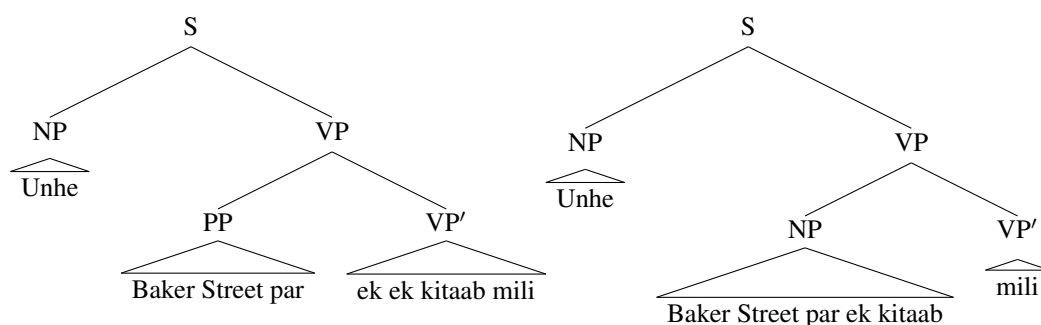
Ambiguity is a pain for every linguistics system , no matter the objective or the language.

Attachment ambiguity is caused due to multiple grammatical function of preposition as follows:

Sita ne khaana khaate hue ek ladke ko dekha.



Unhe Baker Street par ek kitaab mili.



Both these sentences have different syntactic meanings depending on the attachment of the chunks to the head, as can be seen from the trees. The problem in guessing the correct meaning for such constructions is almost impossible, even for humans and only contextual information can help. Taggers show a predictable nature based on data and mark the chunks based on the postpositional marker. For instance, in the second example Baker Street par will be marked as a prepositional phrase almost always.

Coordination ambiguity is also known as structural or syntactic ambiguity. However the present study restricts the term with reference to ambiguity at phrase level.

sabha mein budhe sher aur haathi aaye the

The above sentence can have two interpretations based on the way the coordinate conjunction and its grouping of its arguments.

sabha mein [budhe sher] aur haathi aaye the
sabha mein [budhe sher aur haathi] aaye the

Again, this is hard to resolve even for a human annotator. Taggers usually associated the modifier with the first head alone, again because such data is more common.

(f) Ellipsis

Ellipsis means leaving out words rather than repeating them unnecessarily. While its easy enough for a human to understand such constructions because they can guess the missing parts because of context, its very tough for a tagger to understand what the missing bits refer to. Taggers struggle even with Anaphora resolution where a different marker refers to the entity in question. In the case of ellipsis, the marker is a zero/null marker.

Shyam ne Mohan ko maara par Ram ne nahi
Main baazar gaya, kapde khareede, fer ghar waapas aagya

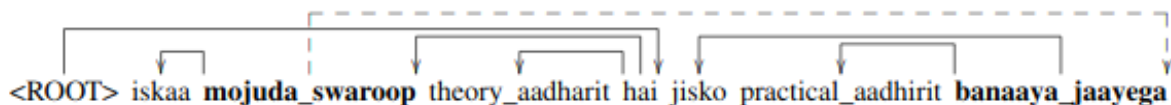
The second examples is that of zero anaphor.

(g) Non-projectivity

Non-projective structures in contrast to projective dependency structures contain a node with a discontinuous yield. These occur pretty frequently in Hindi, especially in longer sentences because of free order. It should be noted that intra-chunk dependencies are projective; all the non-projective edges are distributed among the inter-chunk relations. I spent some time and could find 3 kinds of constructions showing non-projectivity.

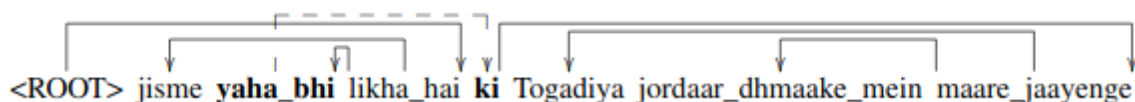
- **Relative Clause Constructions :** Since extraposed relative clauses are separated from the head noun, this dislocation generates discontinuity in the structure.

iskaa mojuda swaroop theory aadharit hai jisko practical aadhirit banaaya jaayega.



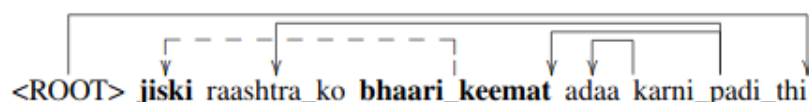
- **Clausal Complement Constructions :** If the head predicate licensing the clausal complement is other than the verb, the canonical order is such that the head is positioned preverbally and its complement is extraposed. In such order the structure has inherent discontinuity.

jisme yaha bhi likha hai ki Togadiya jordaar dhmaake mein maare jaayenge.



- **Genitive Constructions :** In genitive constructions, the genitive marked nominal is easily dislocated from the head noun. Genitive marked noun can either be extraposed or be extracted towards the left.

jiski raashtra ko bhaari keemat adaa karni padi.



As it is evident, in all cases the the arrows between head and dependant cross making it hard for the tagger to find the correct dependant. In a lot of cases the case markers are absent which makes it even harder. One other observation was that marking AnnCorra was often harder in these cases because mutlitple relations seemed to fit for one chunk but it didnt satisfy the others which led to a mismatch.

(h) Particles

Particles are generally attached to a particular word in a sentences to mark emphasis, or contrast. The main particles used in Hindi are: bhi:, hi:, to, tak, bhar, ma:tra.

- **bhi:**

Gita bhi aayi:
Gita ko bhi aana tha

bhi: immediately follows a noun in the direct case and the postposition in the oblique case. The particle bhi: can be used with all types of direct and oblique personal, demonstrative, indefinite, relative, and reflexive pronouns.

tum ho bhi ki nahi?
uska chillaana bhi: thi:k nahi tha:
Ram a:ya: bhi: nahi: Ram bhi: a:ya: nahi:

bhi: occurs b/w main and auxiliary verb. The particle bhi: can be used with conjunct verbs. It is used either between the main verb and the operator (auxiliary verb) or following the main verb and the operator. Note the change in meaning in the last two constructions by moving the particle bhi:

vah laRki: sundar bhi: hE or buddhimaan bhi:
vahan bhi: garmi hE

bhi: always follows the adjective and a similar trend is also seen with adverbs.

aur bhi: sasta:
Ram ko chitr bikul bhi: sundar nahi laga:

bhi: used with "aur" indicates the meaning of 'more'.

- **hi:**

keval Gita hi: aayegi:
Gita ko sahi: jaante hain

hi: can be used as an emphatic marker with nouns. It can also combine with some words as in the second example

puli:s ke aate hi: chor bhaag gaye
puli:s aa hi: rahi thi:

hi: takes different meanings here. In sentence 1 it refers to 'as soon as' while in sentence 2 it's 'just'.

- **to**

Ram a:ya to hE, a:ge dekhte hE(n)
Ram to nahi(n) a:ya

to is mostly used as an emphatic marker and also denotes contrast.

- **tak**

Ram a:ya tak nahi(n)
Ram kal tak pakka: a: ja:ega:

tak takes meanings 'even' or 'up to' as in the above 2 sentences.

- **bhar**

Ram ko muthi bhar ana:j bhi nahi(n) mila
desh bhar me(n) corona fEla hE

bhar takes the meaning of a measurement in both cases. In second it means 'the entire' while in first the unit is mentioned before the particle bhar.

- **ma:tr**

Mujhe ma:tr do minat di:ji:e
a:p kita:b ma:tr de di:ji:e

ma:tr means 'only' as shown in the examples.

As seen in all these cases, the same particle can play multiple roles. This can lead to some ambiguities. However, in most cases it's pretty easy to handle these particles based on the words immediately preceding or following the particle

3 Tag Statistics

3.1 Tag and Markers

3.1.1 UD

- For tag **det**:
2625 instances of ”; 1947 instances of 'is'; 241 instances of 'e'; 126 instances of 'isI'; 373 instances of 'in'; 8 instances of 'inIM'; 158 instances of 'I'; 53 instances of 'isa'; 132 instances of 'us'; 12 instances of 'BI'; 61 instances of 'un'; 1 instances of 'Are'; 28 instances of 'usI'; 27 instances of 'ina'; 1 instances of 'wA'; 1 instances of 'ZI'; 9 instances of 'Z'; 2 instances of 'oM'; 4 instances of 'o'; 1 instances of ','; 8 instances of 'unIM'; 8 instances of 'i'; 1 instances of 'EsI'; 1 instances of 'hIM'; 1 instances of 'ZZ'; 1 instances of 'Ze'; 2 instances of 'aBI';
- For tag **nmod**:
19232 instances of ”; 2061 instances of 'oM'; 131 instances of 'iskA'; 17 instances of 'isameM'; 372 instances of 'I'; 27 instances of 's'; 1538 instances of 'ne'; 281 instances of 'unke'; 106 instances of 'iskI'; 10 instances of 'inameM'; 115 instances of 'iske'; 1325 instances of 'e'; 178 instances of 'M'; 2 instances of 'jI'; 2 instances of ','; 12 instances of 'kara'; 103 instances of 'uske'; 245 instances of 'unkI'; 7 instances of 'kA'; 359 instances of 'iyom'; 55 instances of 'uskA'; 12 instances of 'kI'; 46 instances of 'inmeM'; 11 instances of 'ke'; 7 instances of 'BI'; 22 instances of 'inkI'; 11 instances of 'inke'; 117 instances of 'Z'; 2 instances of 'm'; 19 instances of 'isakA'; 37 instances of 'ArI'; 10 instances of 'isakI'; 48 instances of 'ismeM'; 35 instances of 'Are'; 20 instances of 'ZoM'; 1 instances of 'iyAz'; 35 instances of 'yoM'; 13 instances of 'inkA'; 10 instances of 'Ms'; 90 instances of 'uskI'; 77 instances of 'unkA'; 5 instances of 'isI'; 4 instances of 'nA'; 7 instances of 'uoM'; 10 instances of 'Ziyom'; 49 instances of 'Zne'; 52 instances of 'is'; 7 instances of 'eM'; 1 instances of '333602'; 7 instances of 'us'; 17 instances of 'noM'; 14 instances of 'ArA'; 1 instances of 'riM'; 2 instances of 'se'; 1 instances of 'wA'; 10 instances of 'erI'; 34 instances of 'ie'; 7 instances of 'erA'; 1 instances of 'pawni'; 1 instances of 'uneM'; 15 instances of 'Ze'; 9 instances of 'ja'; 4 instances of 'isa'; 13 instances of 'isse'; 3 instances of 'Ane'; 4 instances of 'a'; 12 instances of 'usse'; 1 instances of 'aI'; 38 instances of 'o'; 11 instances of 'ere'; 1 instances of 'ane'; 1 instances of 'A'; 8 instances of 'isake'; 2 instances of 'unse'; 5 instances of 'inakI'; 1 instances of 'iyA'; 1 instances of 'I'; 5 instances of 'ye'; 1 instances of '()'; 6 instances of 'ZZ'; 4 instances of 'n'; 1 instances of 'Ah'; 14 instances of 'i'; 1 instances of 'Put'; 1 instances of 'hiyom'; 9 instances of 'inake'; 2 instances of 'O'; 4 instances of 'usI'; 1 instances of ','; 5 instances of 'Me'; 2 instances of 'iyAM'; 10 instances of 'unmeM'; 2 instances of 'vAne'; 1 instances of 'bord'; 3 instances of 'inakA'; 1 instances of 'egA'; 1 instances of 'egI'; 1 instances of 'iye'; 1 instances of 'iy'; 1 instances of 'nane'; 1 instances of 'ioM'; 1 instances of '-'; 1 instances of 'ej'; 3 instances of 'un'; 1 instances of 'OY'; 1 instances of 'N'; 1 instances of 'u'; 2 instances of 'yA'; 1 instances of 'unIM'; 1 instances of 'ina'; 1 instances of 'gA'; 4 instances of 'IM'; 1 instances of 'isase'; 1 instances of 'jZa';
- For tag **case**:
5600 instances of 'I'; 41856 instances of ”; 6434 instances of 'e'; 123 instances of 'is'; 11 instances of 'oM'; 2 instances of 'A'; 1 instances of 'wA'; 2 instances of 'o'; 1 instances of 'ke'; 45 instances of 'Z'; 1 instances of 'se'; 1 instances of 'i'; 6 instances of ','; 1 instances of 'yoM'; 3 instances of 'O'; 1 instances of 'M'; 9 instances of 'ZZ'; 1 instances of 'Ze';
- For tag **advmod**:
3105 instances of ”; 18 instances of 'M'; 13 instances of 'I'; 10 instances of 'e'; 29 instances of 'ne'; 5 instances of 'we'; 3 instances of 'Zkara'; 17 instances of 'kara'; 2 instances of ','; 13 instances of 'Z'; 1 instances of 'Ze'; 1 instances of 'A'; 2 instances of 'ie'; 4 instances of 'oM'; 1 instances of 'O'; 1 instances of 'noM'; 4 instances of 'iw'; 1 instances of 'Zne'; 1 instances of 'mAwra';
- For tag **amod**:
716 instances of 'I'; 8910 instances of ”; 460 instances of 'ne'; 1011 instances of 'e'; 76 instances of 'Z'; 11 instances of 'M'; 8 instances of 'kara'; 1 instances of 'maka'; 41 instances of 'ZI'; 50 instances of 'Ze'; 2 instances of 'n'; 8 instances of 'i'; 10 instances of 'IM'; 2 instances of 'nevAll'; 4 instances of 'we'; 9 instances of 'Zne'; 16 instances of 'ie'; 1 instances of 'Ane'; 2 instances of 'ZU'; 1 instances of 'vIM'; 1 instances of 'Me'; 1 instances of 'ul'; 1 instances of 'me'; 1 instances of 'musli'; 17 instances of 'eM'; 4 instances of 'A'; 6 instances of 'is'; 5 instances of 'o'; 1 instances of 'ZyA'; 1 instances of 'nevAle'; 1 instances of 'isa'; 2 instances of 'mI'; 1 instances of 'Zkara'; 2 instances of 'yI'; 1 instances of 'm'; 1 instances of 'MI'; 1 instances of 'nirmiwa'; 1 instances of 'ww'; 1 instances of 'nI'; 1 instances of 'z'; 1 instances of 'ZA'; 2 instances of 'yA'; 1 instances of 'iy'; 1 instances of 'iyA'; 1 instances of 'EM'; 1 instances of 'wI'; 1 instances of 'nA'; 1 instances of 'ue'; 1 instances of 'r';

- For tag **root**:

4270 instances of '"; 962 instances of 'yA'; 41 instances of 'eM'; 163 instances of 'wA'; 1148 instances of 'iyA'; 2689 instances of 'A'; 479 instances of 'e'; 238 instances of 'we'; 170 instances of 'wI'; 12 instances of 'ez'; 56 instances of 'M'; 4 instances of 'wIM'; 18 instances of 'vAyA'; 1 instances of 'yAz'; 1394 instances of 'I'; 82 instances of 'ie'; 1 instances of 'se'; 6 instances of 'iye'; 4 instances of 'iyAz'; 1 instances of 'Iji'; 5 instances of 'kara'; 5 instances of 'iegA'; 82 instances of 'gA'; 101 instances of 'ne'; 134 instances of 'eMge'; 178 instances of 'nA'; 72 instances of 'egI'; 50 instances of 'Mge'; 2 instances of 's'; 97 instances of 'egA'; 24 instances of 'ZA'; 114 instances of 'uA'; 34 instances of 'IM'; 78 instances of 'Z'; 9 instances of 'ZI'; 1 instances of 'vAwA'; 167 instances of 'uI'; 1 instances of 'yAM'; 10 instances of 'oM'; 65 instances of 'i'; 1 instances of 'Ijie'; 69 instances of 'nI'; 5 instances of 'ZegA'; 6 instances of 'ye'; 14 instances of 'Ze'; 23 instances of 'gae'; 3 instances of 'ZyA'; 32 instances of 'ue'; 73 instances of 'gI'; 10 instances of 'gay'; 2 instances of 'UMgI'; 1 instances of 'eMgIM'; 1 instances of 'ZeM'; 3 instances of 'vA'; 1 instances of 'AnA'; 7 instances of 'yI'; 2 instances of 'ZwI'; 4 instances of 'Zne'; 6 instances of 'gaI'; 1 instances of 'ZoM'; 1 instances of 'MnA'; 2 instances of 'iske'; 5 instances of 'ZwA'; 1 instances of 'Zwe'; 1 instances of 'UMgA'; 3 instances of 'unke'; 17 instances of 'eMgI'; 1 instances of 'ke'; 1 instances of 'ZZ'; 10 instances of 'MgI'; 2 instances of 'iy'; 1 instances of 'ZegI'; 1 instances of 'vAwI'; 1 instances of 'iyAM'; 1 instances of 'MA'; 1 instances of 'MvAe'; 1 instances of 'MvAI'; 2 instances of 'iyom'; 2 instances of 'MI'; 2 instances of 'o'; 2 instances of 'Me'; 2 instances of 'UM'; 1 instances of 'noM'; 2 instances of 'uIM'; 1 instances of 'isse'; 1 instances of 'Mne'; 1 instances of 'gIM'; 1 instances of 'zI'; 1 instances of 'egeM'; 1 instances of 'gaIM'; 1 instances of 'MgeM'; 1 instances of 'av'; 1 instances of 'MyA';

- For tag **cop**:

1921 instances of '"; 536 instances of 'M'; 136 instances of 'e'; 114 instances of 'I'; 9 instances of 'UM'; 17 instances of 'IM'; 2 instances of 'z'; 1 instances of 'oM';

- For tag **punct**:

15436 instances of '"; 2541 instances of ','; 638 instances of '"; 5 instances of 'e'; 9 instances of '?'; 2 instances of '!'; 10 instances of 'is'; 1 instances of ',000'; 3 instances of 'I'; 2 instances of ')'; 1 instances of 'Ze'; 1 instances of '...'; 1 instances of ',';

- For tag **obj**:

157 instances of 'ise'; 723 instances of 'oM'; 23 instances of 'inheM'; 545 instances of 'e'; 9014 instances of '"; 135 instances of 'iyom'; 180 instances of 'eM'; 91 instances of 'yA'; 6 instances of 'noM'; 87 instances of 'we'; 286 instances of 'nA'; 28 instances of 'yom'; 2 instances of 's'; 176 instances of 'ne'; 3 instances of 'isase'; 4 instances of 'iyAz'; 2 instances of 'vAyA'; 61 instances of 'wI'; 62 instances of 'egA'; 1 instances of 'inse'; 3 instances of 'ez'; 3 instances of 'ZoM'; 101 instances of 'Z'; 1 instances of 'OoM'; 37 instances of 'gI'; 94 instances of 'eMge'; 319 instances of 'I'; 197 instances of 'A'; 37 instances of 'uA'; 87 instances of 'use'; 66 instances of 'egI'; 21 instances of 'Mge'; 9 instances of 'ue'; 122 instances of 'uneM'; 43 instances of 'iyAM'; 1 instances of 'ZyA'; 323 instances of 'iyA'; 9 instances of 'ZI'; 76 instances of 'M'; 72 instances of 'wA'; 9 instances of 'ZZ'; 10 instances of 'ZnA'; 35 instances of 'uI'; 14 instances of 'i'; 1 instances of 'un'; 1 instances of '0'; 6 instances of 'ZegA'; 44 instances of 'nI'; 1 instances of 'ZeM'; 1 instances of 'Zye'; 11 instances of 'usse'; 10 instances of 'uJe'; 2 instances of 'UMgI'; 6 instances of 'isse'; 2 instances of 'IM'; 42 instances of 'gA'; 4 instances of 'Zne'; 6 instances of 'ZA'; 26 instances of 'ie'; 1 instances of ','; 4 instances of 'gay'; 1 instances of 'y'; 3 instances of 'unko'; 5 instances of 'Ziyom'; 8 instances of 'BI'; 11 instances of 'eMgI'; 13 instances of 'unse'; 13 instances of 'yAM'; 1 instances of 'iske'; 8 instances of 'isko'; 2 instances of 'ZegI'; 6 instances of 'o'; 1 instances of 'ZwA'; 11 instances of 'Ze'; 14 instances of 'is'; 4 instances of 'isI'; 19 instances of 'ineM'; 1 instances of 'ZeMge'; 2 instances of 'wIM'; 1 instances of 'MgI'; 1 instances of 'uJase'; 2 instances of 'vAnA'; 1 instances of 'ZIM'; 1 instances of 'unkI'; 1 instances of 'ke'; 2 instances of 'unke'; 1 instances of 'eg'; 1 instances of 'prwbxXa'; 1 instances of 'OY'; 3 instances of 'uske'; 2 instances of 'ko'; 1 instances of 'gae'; 2 instances of 'isak'; 2 instances of 'ZZiyom'; 1 instances of 'yAz'; 1 instances of 'Ms'; 1 instances of 'UM'; 1 instances of 'uom'; 2 instances of 'kara'; 1 instances of 'yIM'; 1 instances of 'iye'; 1 instances of 'yI'; 1 instances of 'ye'; 1 instances of 'Are'; 2 instances of 'UMgA'; 1 instances of 'uMgI'; 1 instances of 'se'; 1 instances of 'usko'; 1 instances of 'inko'; 1 instances of 'n'; 1 instances of 'iMyom';

- For tag **compound**:

28206 instances of '"; 122 instances of 'I'; 5 instances of 'oM'; 18 instances of 'IM'; 124 instances of 'e'; 9 instances of 'nA'; 1 instances of 'la'; 104 instances of 'M'; 26 instances of 's'; 2 instances of 'sa'; 153 instances of 'Z'; 1 instances of 'iyAz'; 30 instances of 'eM'; 1 instances of 'kara'; 1 instances of 'iyA'; 2 instances of 'wA'; 20 instances of ','; 1 instances of 'e.'; 1 instances of 'ZI'; 9 instances of 'Ze'; 8 instances of 'ZZ'; 1 instances of ',000'; 22 instances of 'n'; 1 instances of '-'; 9 instances of 'iyAM'; 1 instances of 'ZnA'; 3 instances of 'ue'; 5 instances of 'o'; 2 instances of 'uI'; 2 instances of 'Ms'; 6 instances of 'we'; 4 instances of 'ie'; 1 instances of 'egA'; 1

instances of 'r'; 4 instances of 'ne'; 1 instances of 'in'; 5 instances of 'ja'; 1 instances of 'ZA'; 1 instances of 'rrsa'; 2 instances of 'ej'; 1 instances of '329'; 1 instances of 'i'; 3 instances of 'OY'; 3 instances of 'm'; 1 instances of 'S'; 2 instances of 'es'; 1 instances of 'va'; 2 instances of 'A'; 1 instances of 'Me'; 1 instances of 'aX'; 1 instances of 'XaX'; 1 instances of 'ejZ';

- For tag **nsubj**:

12559 instances of ''; 443 instances of 'e'; 9 instances of 'iyAz'; 245 instances of 'eM'; 811 instances of 'oM'; 466 instances of 'nA'; 6 instances of 'ez'; 31 instances of 'M'; 85 instances of 'ne'; 15 instances of 'ko'; 251 instances of 'iyom'; 29 instances of 'iyA'; 4 instances of 'ineM'; 1198 instances of 'unoMne'; 69 instances of 'is'; 2 instances of 'm'; 1 instances of 'ya'; 2 instances of 'yAz'; 116 instances of 'uneM'; 2 instances of 'isakA'; 1 instances of 'isameM'; 1 instances of 'ija'; 1 instances of 'ue'; 31 instances of 'use'; 23 instances of 'A'; 103 instances of 'usne'; 10 instances of 'inhMne'; 5 instances of 'wI'; 76 instances of 'Z'; 2 instances of 'ZeM'; 3 instances of 'uI'; 48 instances of 'I'; 67 instances of 'iyAM'; 12 instances of 'isane'; 5 instances of 'uskA'; 2 instances of 's'; 4 instances of 'inoMne'; 1 instances of 'Mne'; 14 instances of 'BI'; 2 instances of 'Ms'; 1 instances of 'ZoM'; 1 instances of ' '; 17 instances of 'uJe'; 24 instances of 'noM'; 1 instances of 'ZyA'; 4 instances of 'egI'; 2 instances of 'o'; 4 instances of 'isae'; 1 instances of 'eMge'; 50 instances of 'unkA'; 4 instances of 'ie'; 15 instances of 'yA'; 3 instances of 'unse'; 1 instances of 'ZnI'; 2 instances of 'inheM'; 2 instances of 'usI'; 5 instances of 'n'; 5 instances of 'yAM'; 5 instances of 'nI'; 8 instances of 'yoM'; 9 instances of 'ise'; 2 instances of 'ArA'; 2 instances of 'gI'; 1 instances of 'ZnA'; 1 instances of 'ke'; 3 instances of 'Ze'; 2 instances of 'inkA'; 1 instances of 'inko'; 1 instances of 'vAnA'; 3 instances of 'we'; 1 instances of 'gA'; 1 instances of 'iM'; 1 instances of 'ZiyAz'; 1 instances of 'ZA'; 13 instances of 'unke'; 1 instances of 'inase'; 2 instances of 'unkI'; 7 instances of 'unko'; 1 instances of 'isse'; 1 instances of 'isake'; 4 instances of 'uA'; 2 instances of 'inakA'; 3 instances of 'Ziyom'; 2 instances of 'iske'; 1 instances of 'uom'; 2 instances of 'ja'; 1 instances of 'i'; 1 instances of 'd'; 1 instances of 'unoneM'; 11 instances of 'ZZ'; 1 instances of 'erA'; 1 instances of 'unmeM'; 3 instances of 'isne'; 1 instances of 'usse'; 2 instances of 'wA'; 1 instances of 'inmeM'; 1 instances of 'erI'; 1 instances of 'umane'; 1 instances of 'ze'; 1 instances of 'uskI'; 1 instances of 'haa';

- For tag **aux**:

2883 instances of ''; 201 instances of 'wA'; 61 instances of 'uA'; 199 instances of 'we'; 628 instances of 'M'; 645 instances of 'ue'; 290 instances of 'gay'; 1105 instances of 'I'; 37 instances of 'i'; 38 instances of 'uI'; 158 instances of 'wI'; 410 instances of 'A'; 525 instances of 'iyA'; 873 instances of 'e'; 32 instances of 'Mge'; 84 instances of 'gA'; 3 instances of 'Iji'; 1 instances of 'ez'; 242 instances of 'gae'; 28 instances of 'nA'; 1 instances of 'ezge'; 82 instances of 'egA'; 55 instances of 'egI'; 20 instances of 'gI'; 262 instances of 'gaI'; 51 instances of 'IM'; 8 instances of 'gaIM'; 19 instances of 'UM'; 55 instances of 'eMge'; 12 instances of 'gayI'; 19 instances of 'eM'; 113 instances of 'ne'; 1 instances of 'yegA'; 2 instances of 'UMgI'; 39 instances of 'yA'; 12 instances of 'eMgI'; 4 instances of 'kara'; 1 instances of 'Z'; 5 instances of 'nI'; 3 instances of 'ye'; 4 instances of 'wIM'; 2 instances of 'eMgIM'; 1 instances of 'Uz'; 3 instances of 'MgI'; 2 instances of 'gaye'; 2 instances of 'ie'; 1 instances of 'Zwe'; 2 instances of 'yI'; 1 instances of 'MgeM'; 1 instances of 'yeMgI'; 1 instances of 'oM'; 1 instances of 'egeM'; 1 instances of 'o'; 1 instances of 'jAegA';

- For tag **nummod**:

3866 instances of ''; 2 instances of 'n'; 138 instances of 'Z'; 69 instances of 'oM'; 29 instances of 'ZoM'; 310 instances of 'noM'; 1 instances of 'iyom'; 6 instances of '0'; 6 instances of ' ,'; 34 instances of ' ,'; 4 instances of 'e'; 1 instances of ',480'; 3 instances of ',000'; 1 instances of 'o'; 2 instances of 'wIn'; 1 instances of 'M'; 1 instances of ',898';

- For tag **cc**:

5167 instances of ''; 31 instances of ' '; 16 instances of 'is'; 1 instances of 'isse'; 4 instances of 'Z'; 1 instances of 'ina'; 1 instances of 'I'; 1 instances of 'e';

- For tag **obl**:

17975 instances of ''; 1634 instances of 'e'; 96 instances of 'Z'; 12 instances of 's'; 2123 instances of 'oM'; 217 instances of 'isse'; 260 instances of 'iyom'; 17 instances of 'inameM'; 806 instances of 'ne'; 41 instances of 'isake'; 81 instances of 'isameM'; 76 instances of 'unke'; 198 instances of 'is'; 8 instances of 'Are'; 5 instances of 'ke'; 8 instances of 'Ziyom'; 14 instances of 'I'; 24 instances of 'us'; 427 instances of 'iske'; 1 instances of 'oz'; 18 instances of 'we'; 4 instances of 'Ms'; 5 instances of 'ja'; 40 instances of 'o'; 1 instances of 'xillI'; 104 instances of 'ismeM'; 1 instances of 'na'; 8 instances of 'BI'; 2 instances of 'inIM'; 1 instances of 'kara'; 17 instances of 'ZoM'; 439 instances of 'M'; 64 instances of 'uske'; 4 instances of 'nA'; 62 instances of 'isase'; 15 instances of 'isI'; 20 instances of 'usmeM'; 6 instances of 'uom'; 19 instances of 'un'; 39 instances of 'yoM'; 1 instances of 'Ane'; 10 instances of 'IM'; 11 instances of 'Ze'; 4 instances of 'Mne'; 17 instances of 'ie'; 16 instances of 'i'; 15

instances of 'unse'; 11 instances of 'unmeM'; 6 instances of 'inke'; 5 instances of 'unkI'; 20 instances of 'Zne'; 5 instances of 'ina'; 7 instances of 'noM'; 3 instances of 'inse'; 28 instances of 'inmeM'; 2 instances of 'A'; 1 instances of 'vAne'; 16 instances of 'usse'; 2 instances of 'usI'; 1 instances of 'isko'; 3 instances of 'n'; 1 instances of 'jAnabUJaara'; 1 instances of 'rAMj'; 7 instances of 'in'; 8 instances of 'ZZ'; 4 instances of 'm'; 1 instances of 'U'; 2 instances of 'ere'; 9 instances of 'eM'; 5 instances of 'iskI'; 11 instances of 'isa'; 2 instances of 'use'; 1 instances of 'uneM'; 1 instances of 'hAre'; 1 instances of 'MsoM'; 7 instances of 'inake'; 1 instances of 'nI'; 3 instances of 'se'; 5 instances of 'inase'; 1 instances of '0'; 4 instances of 'Me'; 2 instances of 'iyAM'; 1 instances of 'unoMne'; 3 instances of 'isakI'; 1 instances of 'iye'; 1 instances of 'ArA'; 1 instances of 'ZZiyOM'; 4 instances of 'uJa'; 1 instances of 'unkA'; 1 instances of 'ue'; 1 instances of 'inakI'; 1 instances of 'iyA'; 1 instances of 'ej'; 1 instances of 'eMge'; 1 instances of 'zvo'; 1 instances of 'meM'; 1 instances of 'uskI'; 1 instances of 'Aw'; 1 instances of 'uI';

- For tag **advcl**:

1146 instances of 'ne'; 59 instances of 'ke'; 772 instances of 'kara'; 45 instances of 'A'; 632 instances of 'we'; 16 instances of 'i'; 33 instances of 'wI'; 478 instances of '"; 14 instances of 'yA'; 5 instances of 'egA'; 14 instances of 'eM'; 36 instances of 'wA'; 1 instances of 'ez'; 1 instances of 'ue...'; 50 instances of 'I'; 80 instances of 'e'; 6 instances of 'Zwe'; 32 instances of 'Zkara'; 13 instances of 'Z'; 11 instances of 'uA'; 21 instances of 'nA'; 2 instances of 'ZA'; 17 instances of 'ie'; 3 instances of 'Ane'; 3 instances of 'eMge'; 46 instances of 'iyA'; 1 instances of 'gl'; 36 instances of 'Zne'; 7 instances of 'is'; 3 instances of 'ue'; 1 instances of 'ZwI'; 1 instances of 'ZyA'; 5 instances of 'M'; 2 instances of 'Mne'; 4 instances of 'nI'; 1 instances of 'vAI'; 5 instances of 'uI'; 1 instances of 'akara'; 1 instances of 'Zie'; 1 instances of 'iye'; 1 instances of 'uIM'; 3 instances of 'IM'; 2 instances of 'vAkara'; 1 instances of 'UMgA'; 1 instances of 'gae'; 1 instances of 'gA'; 1 instances of 'Mge'; 1 instances of 'Ze'; 1 instances of 'ZwA'; 1 instances of 'kra';

- For tag **mark**:

9758 instances of '"; 1100 instances of 'I'; 677 instances of 'e'; 39 instances of 'oM'; 32 instances of 'is'; 2 instances of 'i'; 3 instances of 'Z'; 1 instances of 'k'; 1 instances of 'nA'; 1 instances of 'we'; 1 instances of '";

- For tag **conj**:

22 instances of 'M'; 4171 instances of '"; 369 instances of 'oM'; 137 instances of 'e'; 3 instances of 'iyAz'; 47 instances of 'we'; 40 instances of 'wA'; 29 instances of 'eM'; 16 instances of 'wI'; 1 instances of 'es'; 129 instances of 'A'; 86 instances of 'iyOM'; 166 instances of 'ne'; 111 instances of 'iyA'; 16 instances of 'ie'; 55 instances of 'yA'; 3 instances of 'ez'; 10 instances of 'yoM'; 169 instances of 'I'; 1 instances of 'yAz'; 47 instances of 'nA'; 16 instances of 'nI'; 21 instances of 'gA'; 8 instances of 'Ze'; 12 instances of 'uA'; 1 instances of 'uoM'; 11 instances of 'uI'; 13 instances of 'gl'; 37 instances of 'Z'; 1 instances of 'riM'; 3 instances of 'unke'; 17 instances of 'eMge'; 5 instances of 'uske'; 2 instances of 'ZnA'; 9 instances of 'egI'; 3 instances of 'Ms'; 18 instances of 'egA'; 10 instances of 'Mge'; 2 instances of 'ZA'; 6 instances of 'i'; 2 instances of 'ZegA'; 1 instances of 'uneM'; 2 instances of 'ZoM'; 8 instances of 'IM'; 2 instances of 'ZwA'; 1 instances of 'MyA'; 1 instances of 'ZyA'; 1 instances of 'Me'; 1 instances of 'ZAnA'; 1 instances of 'r'; 16 instances of 'iyAM'; 3 instances of 'kara'; 5 instances of 'o'; 2 instances of 'Zwe'; 1 instances of 'vA'; 7 instances of 'ue'; 1 instances of 'UMgl'; 1 instances of 'AyA'; 4 instances of 's'; 1 instances of 'Ane'; 3 instances of 'ZZ'; 1 instances of 'gay'; 2 instances of 'gae'; 1 instances of 'UMgA'; 1 instances of 'vAyA'; 1 instances of 'ZI'; 1 instances of 'MnI'; 1 instances of 'ja'; 1 instances of 'vAne'; 1 instances of 'ioM'; 2 instances of 'Zne'; 1 instances of 'UM'; 1 instances of 'unse'; 1 instances of 'iske'; 1 instances of 'unkI';

- For tag **acl**:

770 instances of '"; 448 instances of 'e'; 89 instances of 'A'; 296 instances of 'I'; 1 instances of 'vAI'; 2 instances of 'ise'; 31 instances of 'wI'; 1 instances of 'noM'; 69 instances of 'ie'; 38 instances of 'yA'; 4 instances of 'vAyA'; 47 instances of 'ne'; 7 instances of 'uA'; 103 instances of 'we'; 1 instances of 'ave'; 124 instances of 'ue'; 20 instances of 'nA'; 120 instances of 'uI'; 5 instances of 'Mge'; 29 instances of 'eMge'; 7 instances of 'Zwe'; 34 instances of 'Ze'; 72 instances of 'iyA'; 21 instances of 'egA'; 6 instances of 'ZegA'; 17 instances of 'eM'; 11 instances of 'ZwI'; 9 instances of 'gl'; 12 instances of 'ZI'; 15 instances of 'gA'; 22 instances of 'wA'; 2 instances of 'kara'; 8 instances of 'Me'; 1 instances of 'j'; 11 instances of 'egl'; 10 instances of 'Z'; 11 instances of 'gae'; 1 instances of 'ZyA'; 34 instances of 'i'; 6 instances of 'M'; 2 instances of 'eMgl'; 2 instances of 'ZA'; 1 instances of 'Zegl'; 8 instances of 'nI'; 1 instances of 'iyOM'; 1 instances of 'AyA'; 1 instances of 'ZwA'; 3 instances of 'oM'; 4 instances of 'IM'; 1 instances of 'gay'; 1 instances of 'yl'; 1 instances of 'ye'; 1 instances of 'iy'; 1 instances of 'vAeMge'; 1 instances of 'gal'; 1 instances of 'Zne'; 1 instances of 'MI';

- For tag **dep**:

3873 instances of '"; 8 instances of 'I'; 17 instances of 'e'; 1 instances of 'M'; 20 instances of 'is'; 6 instances of 'Z'; 1 instances of 'A';

- For tag **acl:cl**:
83 instances of 'A'; 3 instances of 'oM'; 334 instances of "; 50 instances of 'e'; 39 instances of 'we'; 18 instances of 'nA'; 26 instances of 'wA'; 8 instances of 'ue'; 104 instances of 'I'; 2 instances of 'IM'; 25 instances of 'wI'; 38 instances of 'yA'; 92 instances of 'iyA'; 9 instances of 'Z'; 7 instances of 'M'; 8 instances of 'ne'; 7 instances of 'gI'; 13 instances of 'uA'; 20 instances of 'uI'; 7 instances of 'egI'; 10 instances of 'gA'; 13 instances of 'eMge'; 1 instances of 'eMgI'; 4 instances of 'nI'; 1 instances of 'ZwI'; 1 instances of 'zA'; 2 instances of 'Ze'; 1 instances of 'ZA'; 1 instances of 'ZegA'; 1 instances of 'ZwA'; 3 instances of 'i'; 5 instances of 'eM'; 3 instances of 'ZI'; 7 instances of 'ie'; 3 instances of 'gae'; 1 instances of 'yI'; 1 instances of 'ere'; 1 instances of 'wIM'; 2 instances of 'Mge'; 7 instances of 'egA'; 1 instances of 'egeM';
- For tag **aux:pass**:
243 instances of 'wA'; 3468 instances of "; 408 instances of 'gae'; 925 instances of 'M'; 978 instances of 'gay'; 30 instances of 'we'; 451 instances of 'gal'; 218 instances of 'iyA'; 428 instances of 'e'; 78 instances of 'wI'; 447 instances of 'I'; 2 instances of 'Mge'; 9 instances of 'gaye'; 311 instances of 'ne'; 111 instances of 'egI'; 240 instances of 'egA'; 217 instances of 'A'; 13 instances of 'yA'; 22 instances of 'gayI'; 30 instances of 'IM'; 1 instances of "; 84 instances of 'nA'; 26 instances of 'nI'; 11 instances of 'gaIM'; 4 instances of 'eMgI'; 10 instances of 'UM'; 33 instances of 'eMge'; 15 instances of 'eM'; 1 instances of 'uA'; 29 instances of 'i'; 3 instances of 'oM'; 3 instances of 'yegA'; 4 instances of 'gA'; 1 instances of 'ie'; 2 instances of 'ya'; 1 instances of 'ye'; 1 instances of 'z'; 1 instances of 'yeMge'; 1 instances of 'wIM'; 1 instances of 'iy'; 1 instances of 'ue';
- For tag **dislocated**:
160 instances of "; 39 instances of 'I'; 66 instances of 'e'; 4 instances of 'is';
- For tag **nsubj:pass**:
362 instances of "; 52 instances of 'oM'; 6 instances of 'eM'; 3 instances of 'I'; 5 instances of 'ise'; 1 instances of 'inameM'; 3 instances of 'iyAM'; 2 instances of 'uske'; 6 instances of 'uneM'; 1 instances of 'Are'; 1 instances of 'iskA'; 14 instances of 'iyOM'; 1 instances of 'inheM'; 1 instances of 'uJe'; 2 instances of 'Z'; 1 instances of 'nA'; 1 instances of 'eMge'; 1 instances of 'isameM'; 2 instances of 'unke'; 6 instances of 'e'; 1 instances of 'BI'; 2 instances of 'ne'; 1 instances of 'yA'; 1 instances of 'A'; 1 instances of 'inke'; 1 instances of 'uI'; 1 instances of 'unkI';
- For tag **xcomp**:
571 instances of "; 12 instances of 'e'; 1 instances of 'n'; 17 instances of 'I'; 1 instances of 'kara'; 1 instances of 'inhMne'; 3 instances of 'nA'; 1 instances of 'eMge'; 8 instances of 'Z'; 1 instances of 'ne'; 1 instances of 'oM'; 3 instances of 'noM'; 1 instances of 'm'; 1 instances of 'IM'; 1 instances of 'Mge';
- For tag **iobj**:
1122 instances of "; 5 instances of 'ko'; 449 instances of 'oM'; 50 instances of 'use'; 13 instances of 'ise'; 1 instances of 'BI'; 7 instances of 'ne'; 96 instances of 'iyOM'; 123 instances of 'uneM'; 13 instances of 'uJe'; 10 instances of 'is'; 26 instances of 'e'; 4 instances of 'noM'; 11 instances of 'yoM'; 1 instances of 'Ms'; 2 instances of 'ja'; 10 instances of 'unse'; 3 instances of 'ZiyOM'; 5 instances of 'inheM'; 3 instances of 'ie'; 7 instances of 'ineM'; 4 instances of 'usse'; 12 instances of 'eM'; 1 instances of 'Zne'; 3 instances of 'Z'; 1 instances of 'se'; 1 instances of 'ZoM'; 3 instances of 'unko'; 1 instances of 'isak'; 2 instances of 'M'; 1 instances of 'o'; 1 instances of 's'; 1 instances of 'inak'; 1 instances of 'unke'; 1 instances of 'iyAM'; 1 instances of 'iMyoM';
- For tag **vocative**:
5 instances of "; 1 instances of 'ie';

3.1.2 AnnCorra

- For tag **pof_cn**:
18725 instances of "; 26 instances of 's'; 2 instances of 'A'; 35 instances of 'e'; 32 instances of 'Z'; 57 instances of 'I'; 77 instances of 'M'; 2 instances of 'Ms'; 2 instances of 'in'; 21 instances of ' '; 1 instances of 'erA'; 19 instances of 'IM'; 3 instances of 'OY'; 1 instances of 'S'; 6 instances of 'ja'; 2 instances of 'ZZ'; 3 instances of 'oM'; 6 instances of 'o'; 22 instances of 'n'; 1 instances of 'e'; 1 instances of 'Ze'; 1 instances of 'ZU'; 1 instances of 'egA'; 2 instances of 'es'; 3 instances of 'eM'; 1 instances of 'rrsa'; 1 instances of 'r'; 3 instances of 'ej'; 1 instances of ',000'; 1 instances of 'ejZ'; 1 instances of 'iyAM'; 1 instances of 'va'; 2 instances of 'm'; 2 instances of 'sa'; 1 instances of '329'; 1 instances of 'wA'; 1 instances of 'XaX';
- For tag **r6**:
10611 instances of "; 1461 instances of 'oM'; 275 instances of 'unke'; 1028 instances of 'e'; 288 instances of

'iyoM'; 8 instances of 'ke'; 238 instances of 'unkI'; 353 instances of 'I'; 76 instances of 'iske'; 80 instances of 'iskI'; 910 instances of 'ne'; 97 instances of 'uske'; 83 instances of 'Z'; 10 instances of 'ere'; 33 instances of 'Are'; 21 instances of 's'; 3 instances of 'isI'; 87 instances of 'M'; 39 instances of 'uskA'; 82 instances of 'iskA'; 6 instances of 'inkA'; 2 instances of 'Me'; 3 instances of ','; 15 instances of 'NULL'; 30 instances of 'Zne'; 22 instances of 'inkI'; 68 instances of 'uskI'; 69 instances of 'unkA'; 27 instances of 'o'; 11 instances of 'i'; 7 instances of 'ZiyoM'; 25 instances of 'ie'; 11 instances of 'yoM'; 6 instances of 'Ms'; 8 instances of 'isake'; 1 instances of 'N'; 16 instances of 'ZoM'; 1 instances of 'us'; 7 instances of 'isakI'; 12 instances of 'isakA'; 39 instances of 'Arl'; 16 instances of 'ArA'; 13 instances of 'kI'; 5 instances of 'erA'; 14 instances of 'Ze'; 1 instances of 'egI'; 5 instances of 'is'; 8 instances of 'erI'; 1 instances of 'm'; 8 instances of 'ja'; 10 instances of 'inke'; 4 instances of 'ZZ'; 5 instances of 'BI'; 1 instances of '-'; 5 instances of 'IM'; 1 instances of 'vAne'; 3 instances of 'ye'; 6 instances of 'inake'; 1 instances of 'isko'; 1 instances of 'unse'; 1 instances of 'uhAre'; 4 instances of 'inakI'; 3 instances of 'Ane'; 6 instances of 'inakA'; 8 instances of 'kA'; 1 instances of 'al'; 9 instances of 'uoM'; 10 instances of 'noM'; 1 instances of 'iy'; 1 instances of 'yA'; 1 instances of 'unIM'; 1 instances of 'ioM'; 1 instances of 'OY'; 1 instances of 'eM'; 1 instances of 'hiyoM'; 1 instances of 'n'; 4 instances of 'usI'; 2 instances of 'a'; 1 instances of 'bord'; 1 instances of 'ane'; 1 instances of 'iye'; 1 instances of 'jI';

- For tag **lwg_psp**:

7113 instances of 'I'; 48259 instances of ','; 7520 instances of 'e'; 51 instances of 'oM'; 40 instances of 'Z'; 2 instances of 'o'; 2 instances of 'i'; 1 instances of 'k'; 1 instances of 'Ze'; 1 instances of 'ke'; 1 instances of 'A'; 9 instances of 'ZZ'; 1 instances of 'se';

- For tag **nmod**:

4174 instances of ','; 1 instances of 'gI'; 154 instances of 'e'; 7 instances of 'i'; 10 instances of 'is'; 28 instances of 'M'; 481 instances of 'ne'; 18 instances of 'NULL'; 106 instances of 'oM'; 5 instances of 'isse'; 6 instances of 'inmeM'; 9 instances of 'Zne'; 16 instances of 'Z'; 5 instances of ','; 45 instances of 'I'; 2 instances of 'n'; 13 instances of 'iyoM'; 1 instances of 'inkI'; 7 instances of 'yoM'; 1 instances of 'iske'; 4 instances of 'we'; 1 instances of 'uneM'; 4 instances of 'Ze'; 1 instances of 'uI'; 1 instances of 'nevAle'; 4 instances of 'unkI'; 1 instances of ','; 1 instances of 'iy'; 1 instances of 'ZA'; 1 instances of '333602'; 1 instances of 'mI'; 9 instances of 'kara'; 6 instances of 'o'; 1 instances of 'A'; 1 instances of 'MI'; 2 instances of 'isameM'; 11 instances of 'usse'; 17 instances of 'ie'; 4 instances of 'eM'; 2 instances of 'ZZ'; 2 instances of 's'; 3 instances of 'us'; 1 instances of 'unse'; 2 instances of 'ZoM'; 2 instances of 'unke'; 1 instances of 'nA'; 3 instances of 'ismeM'; 1 instances of 'iyA'; 1 instances of 'ZyA'; 1 instances of 'wI'; 1 instances of 'Ane'; 2 instances of 'yA'; 2 instances of 'Ms'; 2 instances of 'yI'; 1 instances of 'iyAM'; 2 instances of 'iskI'; 2 instances of 'nevAll'; 1 instances of 'isa'; 1 instances of 'ue'; 2 instances of 'uske'; 1 instances of 'erI'; 1 instances of 'Arl'; 1 instances of 'Are'; 2 instances of 'iskA'; 1 instances of 'erA'; 1 instances of 'Zkara';

- For tag **k7**:

5446 instances of ','; 52 instances of 'inmeM'; 818 instances of 'oM'; 93 instances of 'isameM'; 94 instances of 'iyoM'; 447 instances of 'ne'; 935 instances of 'e'; 207 instances of 'is'; 19 instances of 'iske'; 152 instances of 'ismeM'; 50 instances of 'M'; 41 instances of 'Z'; 18 instances of 'uske'; 1 instances of 'I'; 4 instances of 'Me'; 1 instances of 'inke'; 26 instances of 'us'; 30 instances of 'yoM'; 13 instances of 'isa'; 6 instances of 'ina'; 18 instances of 'un'; 1 instances of 'uJame'; 6 instances of 'NULL'; 4 instances of 's'; 14 instances of 'unke'; 1 instances of 'm'; 1 instances of 'Ms'; 8 instances of 'noM'; 18 instances of 'o'; 1 instances of 'we'; 4 instances of 'uJa'; 5 instances of 'Ze'; 8 instances of 'ie'; 12 instances of 'Zne'; 22 instances of 'inameM'; 1 instances of 'usI'; 2 instances of 'isake'; 22 instances of 'unmeM'; 2 instances of 'IM'; 5 instances of 'i'; 1 instances of 'meM'; 21 instances of 'usmeM'; 2 instances of 'ZiyoM'; 3 instances of 'isI'; 1 instances of 'Mne'; 6 instances of 'Are'; 1 instances of 'inIM'; 4 instances of 'isse'; 1 instances of ','; 3 instances of 'n'; 13 instances of 'ZoM'; 5 instances of 'uoM'; 2 instances of 'BI'; 1 instances of 'uskI'; 1 instances of 'ja'; 4 instances of 'in'; 2 instances of 'ZZ'; 3 instances of 'ke'; 1 instances of 'hAre'; 1 instances of 'iye'; 1 instances of 'se'; 1 instances of 'A'; 2 instances of 'inake'; 1 instances of 'Aw'; 1 instances of 'Ane';

- For tag **r6-k2**:

8 instances of 'yoM'; 1171 instances of ','; 14 instances of 'I'; 578 instances of 'ne'; 10 instances of 'ie'; 114 instances of 'e'; 258 instances of 'oM'; 20 instances of 'unkI'; 7 instances of 'M'; 43 instances of 'iskA'; 16 instances of 'uskI'; 5 instances of 'Z'; 12 instances of 'unkA'; 36 instances of 'iyoM'; 3 instances of 'isakI'; 13 instances of 'Zne'; 28 instances of 'iskI'; 4 instances of 'inkA'; 12 instances of 'uskA'; 4 instances of 'i'; 2 instances of 'iske'; 4 instances of 'NULL'; 1 instances of 'pawni'; 2 instances of 'Ze'; 7 instances of 'isakA'; 3 instances of 'o'; 1 instances of 'Ah'; 1 instances of 'kI'; 1 instances of 'BI'; 2 instances of 'uske'; 3 instances of 'unke'; 2 instances of 'noM'; 3 instances of 'is'; 2 instances of 'Me'; 1 instances of 'ye'; 1 instances of 'isI'; 1 instances of 'nane'; 2 instances of 'erI'; 1 instances of 'jI';

- For tag **pof**:

11196 instances of '"; 78 instances of 'e'; 124 instances of 'Z'; 3 instances of 'ne'; 31 instances of 'M'; 62 instances of 'I'; 9 instances of 'Ze'; 7 instances of 'ZZ'; 10 instances of 'nA'; 1 instances of 'm'; 28 instances of 'eM'; 1 instances of 'kara'; 9 instances of 'iyAM'; 1 instances of 'Me'; 11 instances of 'NULL'; 1 instances of 'ZA'; 1 instances of 'la'; 3 instances of 'oM'; 1 instances of 'ZnA'; 2 instances of 'n'; 1 instances of 'iyAz'; 1 instances of 'aX'; 1 instances of 'A'; 1 instances of '-'; 1 instances of 'wA'; 1 instances of 'u';

- For tag **rt**:

870 instances of 'ne'; 142 instances of 'oM'; 791 instances of '"; 102 instances of 'iske'; 21 instances of 'iyom'; 44 instances of 'e'; 6 instances of 'o'; 25 instances of 'Zne'; 6 instances of 'Z'; 1 instances of 'Mne'; 7 instances of 'M'; 1 instances of 'in'; 1 instances of 'usI'; 18 instances of 'unke'; 1 instances of 'is'; 4 instances of 'ke'; 3 instances of 'BI'; 7 instances of 'uske'; 1 instances of 'Ane'; 1 instances of 'ZZ'; 3 instances of 'Are'; 1 instances of 'isse'; 2 instances of 'ere'; 1 instances of 'uoM'; 3 instances of 's'; 1 instances of 'iskI'; 3 instances of 'isake'; 1 instances of 'A'; 2 instances of 'Ziyom'; 1 instances of 'na'; 1 instances of 'yoM'; 1 instances of 'Ze'; 1 instances of 'Ms'; 1 instances of 'kara'; 1 instances of 'noM'; 1 instances of 'ZoM'; 1 instances of 'unkI';

- For tag **ras-pof**:

4 instances of '"; 1 instances of 'e'; 1 instances of 'oM';

- For tag **main**:

1275 instances of 'I'; 1055 instances of 'iyA'; 2685 instances of 'A'; 938 instances of 'yA'; 5309 instances of '"; 478 instances of 'e'; 138 instances of 'wI'; 139 instances of 'wA'; 89 instances of 'NULL'; 219 instances of 'we'; 352 instances of 'M'; 30 instances of 'ue'; 138 instances of 'uI'; 135 instances of 'nA'; 63 instances of 'egI'; 57 instances of 'gI'; 121 instances of 'eMge'; 62 instances of 'i'; 84 instances of 'ne'; 62 instances of 'nI'; 2 instances of 'MI'; 19 instances of 'eM'; 3 instances of 'ZwA'; 12 instances of 'Ze'; 48 instances of 'Z'; 22 instances of 'ZA'; 95 instances of 'uA'; 77 instances of 'egA'; 45 instances of 'Mge'; 38 instances of 'IM'; 2 instances of 'Me'; 1 instances of 'uIM'; 74 instances of 'gA'; 8 instances of 'gay'; 1 instances of 'eMgIM'; 76 instances of 'ie'; 3 instances of 'wIM'; 16 instances of 'eMgI'; 2 instances of 'Zne'; 17 instances of 'vAyA'; 13 instances of 'gae'; 1 instances of 'vAwI'; 1 instances of 'MyA'; 1 instances of 'UMgI'; 4 instances of 'UM'; 4 instances of 'ZegA'; 6 instances of 'ZI'; 1 instances of 'Zwe'; 1 instances of 'UMgA'; 7 instances of '"; 2 instances of 'iy'; 6 instances of 'iye'; 4 instances of 'yI'; 8 instances of 'MgI'; 3 instances of 'ZyA'; 2 instances of 'iegA'; 11 instances of 'ez'; 1 instances of 'ZegI'; 1 instances of 'zI'; 2 instances of 'vA'; 1 instances of 'null'; 1 instances of 'kara'; 1 instances of 'gIM'; 2 instances of 'gal'; 1 instances of 'MgeM'; 1 instances of 'AIM'; 1 instances of 'MA'; 1 instances of 'galM'; 1 instances of 'Mne'; 1 instances of 'MvAe'; 1 instances of 'MvAI'; 1 instances of 'z'; 1 instances of 'vAwA'; 1 instances of 'ZwI'; 1 instances of 'MnA'; 4 instances of 'ye';

- For tag **lwg_vaux**:

3577 instances of '"; 1167 instances of 'gay'; 203 instances of 'wI'; 1258 instances of 'I'; 770 instances of 'iyA'; 233 instances of 'we'; 672 instances of 'M'; 688 instances of 'ue'; 668 instances of 'gal'; 667 instances of 'gae'; 437 instances of 'ne'; 1040 instances of 'e'; 405 instances of 'A'; 17 instances of 'eMgI'; 52 instances of 'IM'; 342 instances of 'wA'; 35 instances of 'eM'; 84 instances of 'eMge'; 33 instances of 'Mge'; 159 instances of 'egI'; 57 instances of 'yA'; 68 instances of 'uA'; 71 instances of 'i'; 9 instances of 'Z'; 8 instances of 'Ze'; 123 instances of 'nA'; 1 instances of 'Zwe'; 88 instances of 'gA'; 18 instances of 'galM'; 33 instances of 'nI'; 287 instances of 'egA'; 39 instances of 'uI'; 34 instances of 'ZA'; 31 instances of 'gayI'; 3 instances of 'ie'; 19 instances of 'UM'; 21 instances of 'gI'; 3 instances of 'UMgI'; 11 instances of 'ZI'; 3 instances of 'yegA'; 4 instances of 'ZegA'; 1 instances of 'oM'; 1 instances of 'ezge'; 5 instances of 'ZwA'; 1 instances of 'iy'; 2 instances of 'eMgIM'; 12 instances of 'gaye'; 1 instances of 'jAegA'; 5 instances of 'wIM'; 1 instances of 'ZegI'; 1 instances of 'ya'; 1 instances of 'ez'; 4 instances of 'kara'; 3 instances of 'MgI'; 1 instances of 'MgeM'; 3 instances of 'Iji'; 1 instances of 'iye'; 2 instances of 'yI'; 3 instances of 'ye'; 2 instances of 'ZwI'; 1 instances of 'egeM'; 1 instances of 'yeMge'; 1 instances of 'yeMgI'; 1 instances of 'o'; 1 instances of 'Uz';

- For tag **lwg_vaux_ont**:

354 instances of 'I'; 3132 instances of '"; 1006 instances of 'M'; 167 instances of 'gay'; 33 instances of 'IM'; 140 instances of 'wA'; 359 instances of 'e'; 226 instances of 'A'; 53 instances of 'egA'; 18 instances of 'we'; 82 instances of 'gal'; 11 instances of 'UM'; 8 instances of 'eMge'; 1 instances of '"; 53 instances of 'wI'; 3 instances of 'oM'; 21 instances of 'gae'; 7 instances of 'ne'; 2 instances of 'galM'; 6 instances of 'gayI'; 1 instances of 'z'; 4 instances of 'gA'; 11 instances of 'egI'; 2 instances of 'Mge'; 1 instances of 'ya'; 1 instances of 'gI'; 1 instances of 'yegA'; 1 instances of 'ue'; 6 instances of 'eM'; 1 instances of 'Ze'; 1 instances of 'ye'; 1 instances of 'uA'; 2 instances of 'nA'; 1 instances of 'eMgI'; 1 instances of 'nI';

- For tag **rsym**:

16271 instances of '"; 3356 instances of ',"; 688 instances of '"'; 10 instances of '??'; 2 instances of '.,'; 1 instances of '...'; 1 instances of ',-'; 2 instances of '!''; 2 instances of '))';

- For tag **k1**:

14718 instances of '"; 1268 instances of 'unoMne'; 499 instances of 'nA'; 482 instances of 'e'; 155 instances of 'NULL'; 812 instances of 'oM'; 265 instances of 'iyoM'; 26 instances of 'I'; 53 instances of 'unkA'; 17 instances of 'unke'; 24 instances of 'noM'; 94 instances of 'ne'; 4 instances of 'ineM'; 71 instances of 'iyAM'; 10 instances of 'inhMne'; 63 instances of 'uneM'; 261 instances of 'eM'; 17 instances of 'is'; 30 instances of 'M'; 113 instances of 'usne'; 9 instances of 'ko'; 4 instances of 'ZiyoM'; 89 instances of 'Z'; 13 instances of 'iyAz'; 24 instances of 'use'; 2 instances of 'usI'; 7 instances of 'ez'; 4 instances of 'inoMne'; 2 instances of 'inkA'; 3 instances of 'o'; 1 instances of 'uI'; 2 instances of 'inakA'; 7 instances of ',"; 12 instances of 'BI'; 13 instances of 'ZZ'; 5 instances of 'uskA'; 8 instances of 'iyA'; 1 instances of 'erI'; 1 instances of 'haa'; 7 instances of 'A'; 1 instances of 'Mne'; 2 instances of 'm'; 3 instances of 'ZeM'; 10 instances of 'ise'; 7 instances of 'yAM'; 2 instances of 'ArA'; 3 instances of 'yAz'; 2 instances of 'isake'; 4 instances of 'isae'; 4 instances of 'uJe'; 2 instances of 'Ms'; 2 instances of 'inheM'; 9 instances of 'yoM'; 2 instances of 'unko'; 1 instances of 'inke'; 1 instances of 'umane'; 3 instances of 'isne'; 1 instances of 'unoneM'; 4 instances of 'unkI'; 1 instances of 'ZoM'; 1 instances of 'usse'; 5 instances of 'Ze'; 1 instances of 'inko'; 13 instances of 'isane'; 2 instances of 'uske'; 3 instances of 'unse'; 2 instances of 'ja'; 5 instances of 'n'; 2 instances of 'i'; 1 instances of 'iM'; 5 instances of 'nI'; 3 instances of 's'; 1 instances of 'ya'; 1 instances of 'vAnA'; 1 instances of 'uoM'; 1 instances of 'isse'; 1 instances of 'erA'; 1 instances of 'ZnI'; 1 instances of 'd'; 1 instances of 'Are'; 1 instances of 'iskA'; 3 instances of 'yA'; 2 instances of 'iske'; 1 instances of 'ZA'; 2 instances of 'isakA'; 2 instances of 'isameM'; 1 instances of 'ze'; 2 instances of 'ZnA'; 1 instances of 'inameM'; 1 instances of 'unmeM'; 1 instances of 'iskI'; 1 instances of 'inase'; 1 instances of 'inmeM'; 1 instances of 'uskI'; 1 instances of 'ZiyAz';

- For tag **k4**:

1210 instances of '"; 468 instances of 'oM'; 27 instances of 'e'; 135 instances of 'uneM'; 99 instances of 'iyoM'; 7 instances of 'ineM'; 1 instances of 'isak'; 50 instances of 'use'; 3 instances of 'unko'; 15 instances of 'uJe'; 6 instances of 'inheM'; 12 instances of 'yoM'; 1 instances of 'uoM'; 12 instances of 'eM'; 5 instances of 'usse'; 2 instances of 'se'; 10 instances of 'unse'; 1 instances of 'BI'; 14 instances of 'ise'; 4 instances of 'noM'; 12 instances of 'is'; 1 instances of 's'; 1 instances of 'unke'; 3 instances of 'ie'; 1 instances of 'iMyoM'; 1 instances of 'M'; 7 instances of 'ko'; 2 instances of 'ja'; 8 instances of 'ne'; 1 instances of 'usko'; 1 instances of 'o'; 1 instances of ',"; 1 instances of 'inak'; 4 instances of 'Z'; 1 instances of 'Ms'; 2 instances of 'NULL'; 1 instances of 'iyAM'; 3 instances of 'ZiyoM'; 1 instances of 'ZoM'; 1 instances of 'Zne';

- For tag **k2**:

11658 instances of '"; 171 instances of 'ise'; 5 instances of ',"; 30 instances of 'yoM'; 446 instances of 'e'; 132 instances of 'iyoM'; 127 instances of 'uneM'; 740 instances of 'oM'; 74 instances of 'Z'; 43 instances of 'NULL'; 1 instances of 'ZiyAz'; 143 instances of 'eM'; 183 instances of 'nA'; 8 instances of 'nI'; 39 instances of 'M'; 4 instances of 'isI'; 2 instances of 'kara'; 4 instances of 'i'; 15 instances of 'I'; 3 instances of 's'; 14 instances of 'unse'; 23 instances of 'inheM'; 13 instances of 'yAM'; 135 instances of 'ne'; 5 instances of 'ZiyoM'; 47 instances of 'iyAM'; 91 instances of 'use'; 12 instances of 'is'; 1 instances of 'yAz'; 20 instances of 'ineM'; 3 instances of 'ez'; 10 instances of 'uJe'; 11 instances of 'usse'; 10 instances of 'isko'; 5 instances of 'o'; 1 instances of 'UM'; 3 instances of 'ie'; 1 instances of 'ZeM'; 3 instances of 'isase'; 1 instances of 'IM'; 1 instances of 'un'; 6 instances of 'noM'; 3 instances of 'egI'; 4 instances of 'ZnA'; 4 instances of 'we'; 3 instances of 'unko'; 8 instances of 'BI'; 1 instances of 'gA'; 3 instances of 'A'; 5 instances of 'iyA'; 1 instances of 'inse'; 1 instances of 'usko'; 1 instances of 'iMyoM'; 6 instances of 'isse'; 3 instances of 'uske'; 1 instances of 'unmeM'; 4 instances of 'iyAz'; 7 instances of 'ZZ'; 2 instances of 'isak'; 2 instances of 'ZoM'; 1 instances of 'unkI'; 4 instances of 'wI'; 1 instances of 'n'; 8 instances of 'Ze'; 2 instances of 'wA'; 3 instances of 'ko'; 3 instances of 'Zne'; 1 instances of 'inko'; 1 instances of 'se'; 1 instances of '0,'; 1 instances of 'gI'; 1 instances of 'uoM'; 1 instances of 'uMgI'; 4 instances of 'yA'; 2 instances of 'vAnA'; 2 instances of 'eMge'; 1 instances of 'Mge'; 1 instances of 'eMgI'; 1 instances of 'unke'; 1 instances of 'ZZiyoM'; 1 instances of 'y'; 1 instances of 'OoM'; 1 instances of 'Mne'; 1 instances of 'uJase'; 1 instances of 'OY'; 1 instances of 'ZI'; 1 instances of 'iske';

- For tag **ccof**:

746 instances of 'iyA'; 11495 instances of '"; 956 instances of 'I'; 209 instances of 'NULL'; 54 instances of 'Mge'; 135 instances of 'eM'; 632 instances of 'oM'; 137 instances of 'egA'; 3 instances of 'iegA'; 283 instances of 'yA'; 105 instances of 'uI'; 4 instances of 'gaI'; 251 instances of 'we'; 8 instances of 'vAyA'; 272 instances of 'nA'; 573 instances of 'e'; 102 instances of 'egI'; 21 instances of 'Ze'; 351 instances of 'M'; 16 instances of 'eMgI'; 395 instances of 'ne'; 178 instances of 'wI'; 105 instances of 'gA'; 211 instances of 'wA'; 101 instances of 'Z';

77 instances of 'nI'; 15 instances of 'gae'; 96 instances of 'uA'; 618 instances of 'A'; 4 instances of 'Me'; 147 instances of 'iyOM'; 29 instances of 'i'; 173 instances of 'eMge'; 9 instances of 'gay'; 24 instances of 'yoM'; 7 instances of 'ZwA'; 10 instances of 'Zne'; 4 instances of 'ZyA'; 30 instances of 'IM'; 18 instances of 'o'; 73 instances of 'ie'; 13 instances of 'ZI'; 9 instances of 's'; 9 instances of 'UM'; 91 instances of 'gl'; 25 instances of 'ue'; 5 instances of 'uske'; 4 instances of 'ZZ'; 22 instances of 'iyAM'; 1 instances of 'use'; 1 instances of 'vAI'; 3 instances of 'vA'; 6 instances of 'Ms'; 1 instances of 'AoM'; 1 instances of 'ZeM'; 3 instances of 'Zwe'; 7 instances of 'unke'; 2 instances of 'N'; 1 instances of 'MnI'; 1 instances of 'ZAnA'; 5 instances of 'yI'; 5 instances of 'UMgl'; 1 instances of 'erA'; 1 instances of 'yAz'; 1 instances of 'w'; 1 instances of 'ZeMge'; 1 instances of 'Ane'; 4 instances of 'UMgA'; 4 instances of 'MgI'; 15 instances of 'ZegA'; 5 instances of 'uneM'; 2 instances of 'Zegl'; 1 instances of 'ioM'; 1 instances of 'vAeMge'; 1 instances of 'iske'; 2 instances of 'vAne'; 1 instances of 'yIM'; 1 instances of 'MyA'; 1 instances of 'Zye'; 5 instances of 'iyAz'; 2 instances of 'unkI'; 3 instances of 'ZwI'; 8 instances of 'kara'; 1 instances of 'egeM'; 2 instances of 'AyA'; 4 instances of ','; 1 instances of 'es'; 8 instances of 'ZnA'; 1 instances of 'AnA'; 3 instances of 'wIM'; 1 instances of 'ZAyA'; 2 instances of 'uoM'; 1 instances of 'ZIM'; 1 instances of 'Zkara'; 2 instances of 'un'; 5 instances of 'ZoM'; 1 instances of 'unse'; 1 instances of 'eg'; 1 instances of 'yAM'; 1 instances of 'av'; 5 instances of 'ez'; 3 instances of 'n'; 1 instances of 'unko'; 12 instances of 'ZA'; 3 instances of 'ja'; 3 instances of 'iye'; 1 instances of 'r'; 1 instances of 'usne'; 2 instances of 'riM'; 4 instances of 'ye'; 3 instances of 'uIM'; 1 instances of 'yU'; 1 instances of 'ke'; 2 instances of 'iy'; 1 instances of 'ZZiyOM'; 1 instances of 'Iji'; 1 instances of 'ue...'; 1 instances of 'z'; 1 instances of 'Ijie'; 1 instances of 'uskI'; 1 instances of 'MvAe'; 1 instances of 'NA'; 1 instances of 'ZiyOM';

- For tag **nmod_adj**:

2021 instances of 'is'; 18419 instances of ''; 27 instances of 'M'; 869 instances of 'I'; 1 instances of 'u'; 321 instances of 'noM'; 1293 instances of 'e'; 389 instances of 'in'; 136 instances of 'us'; 89 instances of 'oM'; 237 instances of 'Z'; 50 instances of 'Ze'; 48 instances of 'ZI'; 134 instances of 'isI'; 59 instances of 'isa'; 30 instances of 'ZoM'; 30 instances of 'usI'; 5 instances of 'o'; 32 instances of 'ina'; 40 instances of ','; 4 instances of ',000'; 19 instances of 'eM'; 61 instances of 'un'; 12 instances of 'BI'; 2 instances of 'wIn'; 8 instances of 'unIM'; 2 instances of 'me'; 11 instances of 'IM'; 1 instances of 'mI'; 1 instances of ',898'; 6 instances of '0'; 1 instances of 'r'; 1 instances of 'z'; 1 instances of 'Me'; 1 instances of 'Ms'; 6 instances of ',,'; 4 instances of 'A'; 2 instances of 'aBI'; 1 instances of 'Put'; 1 instances of 'hIM'; 8 instances of 'inIM'; 4 instances of 'n'; 1 instances of ',480'; 1 instances of 'maka'; 1 instances of 'musli'; 2 instances of 'm'; 1 instances of 'EsI'; 1 instances of 'ZZ'; 1 instances of 'vIM'; 1 instances of 'EM'; 2 instances of 'i'; 1 instances of 'nirmiwa'; 1 instances of 'iyOM'; 1 instances of 's'; 1 instances of 'ZU'; 1 instances of 'ww'; 1 instances of 'Are'; 1 instances of 'a';

- For tag **rs**:

1432 instances of ''; 3 instances of 'egA'; 5 instances of 'I'; 3 instances of 'oM'; 4 instances of 'A'; 1 instances of ','; 5 instances of 'e'; 2 instances of 'iyA'; 1 instances of 'eMge'; 3 instances of 'nA'; 6 instances of 'NULL'; 1 instances of 'ne'; 1 instances of 'ej'; 1 instances of '()'; 1 instances of 'eM'; 1 instances of 'uI'; 1 instances of 'Z'; 1 instances of 'Ze'; 2 instances of 'gA'; 1 instances of 'Zegl'; 2 instances of 'ie'; 1 instances of 'egl'; 1 instances of 'i'; 1 instances of '0'; 1 instances of 'iyAM'; 1 instances of 'yA'; 1 instances of ',';

- For tag **fragof**:

40 instances of 'I'; 166 instances of ''; 70 instances of 'e'; 3 instances of 'is'; 1 instances of 'wI'; 1 instances of 'M';

- For tag **k7t**:

5123 instances of ''; 122 instances of 'iske'; 254 instances of 'e'; 85 instances of 'isse'; 1 instances of 'I'; 4 instances of 'Mne'; 195 instances of 'oM'; 19 instances of 'we'; 12 instances of 'Z'; 324 instances of 'ne'; 7 instances of 'isake'; 28 instances of 'uske'; 8 instances of 'ie'; 23 instances of 'o'; 18 instances of 'NULL'; 10 instances of 'M'; 2 instances of 'Ze'; 3 instances of 'm'; 8 instances of 'iyOM'; 2 instances of 'BI'; 5 instances of 'Zne'; 7 instances of 'i'; 1 instances of 'nA'; 1 instances of 'kara'; 1 instances of 'xillI'; 1 instances of 'ja'; 1 instances of 'uI'; 1 instances of 'ZoM'; 1 instances of 'unse';

- For tag **adv**:

1807 instances of ''; 69 instances of 'e'; 5 instances of 'I'; 22 instances of 'M'; 3 instances of 'Zkara'; 5 instances of 'we'; 7 instances of 'oM'; 15 instances of 'kara'; 13 instances of 'Z'; 1 instances of 'iske'; 4 instances of 'ne'; 6 instances of 'iw'; 1 instances of 'mAwra'; 1 instances of 'NULL'; 1 instances of 'A';

- For tag **k4a**:

71 instances of 'uneM'; 17 instances of 'uJe'; 76 instances of 'oM'; 19 instances of 'eM'; 167 instances of ''; 9 instances of 'is'; 16 instances of 'use'; 7 instances of 'iyOM'; 1 instances of 'ke'; 4 instances of 'ise'; 8 instances

of 'ko'; 2 instances of 'inheM'; 2 instances of 'e'; 5 instances of 'unko'; 1 instances of 'heM'; 1 instances of 'ija'; 3 instances of 'BI'; 1 instances of 'ineM';

- For tag **vmod**:
1391 instances of ''; 628 instances of 'we'; 821 instances of 'kara'; 28 instances of 'iyoM'; 17 instances of 'i'; 1 instances of 'IM'; 8 instances of 'wI'; 9 instances of 'ie'; 116 instances of 'ne'; 72 instances of 'e'; 12 instances of 'Z'; 20 instances of 'A'; 63 instances of 'ke'; 7 instances of 'iyA'; 4 instances of 'uA'; 1 instances of 'Zie'; 2 instances of 'isse'; 77 instances of 'oM'; 128 instances of 'iske'; 18 instances of 'I'; 24 instances of 'unke'; 6 instances of 'Zwe'; 4 instances of 'isake'; 13 instances of 'NULL'; 1 instances of 'ZZ'; 1 instances of 'eMge'; 10 instances of 'M'; 32 instances of 'Zkara'; 9 instances of 'uske'; 12 instances of 'nA'; 3 instances of 'yoM'; 1 instances of 'yA'; 9 instances of 'wA'; 7 instances of 'inke'; 1 instances of 'ZoM'; 3 instances of 'Zne'; 3 instances of 'ue'; 1 instances of 'akara'; 1 instances of 'ZyA'; 2 instances of 'Ane'; 8 instances of 'eM'; 2 instances of 'is'; 2 instances of 'vAkara'; 1 instances of 's'; 1 instances of 'gA'; 1 instances of 'gae'; 3 instances of 'nI'; 1 instances of 'ZI'; 1 instances of 'jAnabUJaara'; 1 instances of 'egA'; 4 instances of 'inake'; 1 instances of 'isI'; 1 instances of 'egeM'; 1 instances of 'kra'; 1 instances of 'ZA'; 1 instances of 'ZiyoM'; 1 instances of 'rAMj';
- For tag **sent-adv**:
68 instances of 'iske'; 5 instances of 'e'; 279 instances of ''; 36 instances of 'M'; 5 instances of 'IM'; 2 instances of 'isI'; 1 instances of 'ne'; 1 instances of 'Z'; 1 instances of 'isase'; 4 instances of 'isse'; 4 instances of 'I'; 1 instances of 'NULL'; 1 instances of 'unke';
- For tag **lwg_rp**:
4144 instances of ''; 12 instances of 'e'; 8 instances of 'I'; 6 instances of 'Z'; 1 instances of 'A';
- For tag **lwg_neg**:
1945 instances of '';
- For tag **k1s**:
2410 instances of ''; 33 instances of 'NULL'; 13 instances of 'ne'; 40 instances of 'I'; 41 instances of 'e'; 1 instances of 'pawni'; 26 instances of 'Z'; 4 instances of 'noM'; 3 instances of 'M'; 12 instances of 'nA'; 1 instances of 'm'; 1 instances of 'prwbxXa'; 1 instances of 'inhMne'; 1 instances of 'iyAM'; 1 instances of 'eM'; 1 instances of 'oM'; 1 instances of '6.8'; 1 instances of 'n'; 1 instances of 'iyA'; 1 instances of 'kara'; 1 instances of 'A'; 1 instances of 'wA'; 1 instances of 'se';
- For tag **mod**:
42 instances of 'e'; 21 instances of 'Z'; 1835 instances of ''; 324 instances of 'is'; 8 instances of ','; 3 instances of 'wA'; 3 instances of '0'; 15 instances of 'I'; 2 instances of 'oM'; 1 instances of 'isse'; 1 instances of 'yoM'; 1 instances of 'Ze'; 1 instances of ',000'; 1 instances of 'iyoM'; 2 instances of 'M'; 1 instances of 'ZU'; 1 instances of 'ina'; 1 instances of 'we'; 1 instances of 'A';
- For tag **k2s**:
384 instances of ''; 3 instances of 'NULL'; 5 instances of 'nA'; 5 instances of 'e'; 3 instances of 'Z'; 1 instances of 'ZI'; 1 instances of 'A'; 1 instances of 'eM'; 2 instances of 'ne'; 2 instances of 'ise'; 1 instances of 'iyA'; 1 instances of 'oM'; 1 instances of 'BI'; 1 instances of 'wA';
- For tag **nmod_relc**:
353 instances of ''; 96 instances of 'I'; 91 instances of 'A'; 27 instances of 'wI'; 8 instances of 'Z'; 62 instances of 'e'; 42 instances of 'yA'; 25 instances of 'wA'; 1 instances of 'wIM'; 14 instances of 'uA'; 92 instances of 'iyA'; 35 instances of 'M'; 1 instances of 'egeM'; 19 instances of 'uI'; 7 instances of 'ue'; 35 instances of 'we'; 2 instances of 'Ze'; 8 instances of 'ie'; 16 instances of 'eMge'; 4 instances of 'nI'; 8 instances of 'gI'; 7 instances of 'egI'; 7 instances of 'ne'; 1 instances of 'yI'; 4 instances of 'gae'; 1 instances of 'ZwI'; 19 instances of 'nA'; 10 instances of 'gA'; 3 instances of 'IM'; 6 instances of 'egA'; 3 instances of 'ZI'; 3 instances of 'eM'; 1 instances of 'ZegA'; 1 instances of 'eMgI'; 1 instances of 'zA'; 1 instances of 'ZA'; 1 instances of 'ZwA'; 1 instances of 'oM'; 1 instances of 'MgI'; 1 instances of 'ez'; 3 instances of 'i'; 3 instances of 'Mge'; 1 instances of 'NULL';
- For tag **k7p**:
3439 instances of ''; 492 instances of 'oM'; 318 instances of 'M'; 242 instances of 'e'; 3 instances of 'ZoM'; 1 instances of 'usmeM'; 22 instances of 'Z'; 2 instances of 'inameM'; 3 instances of 'uske'; 2 instances of 'Are'; 19 instances of 'NULL'; 3 instances of 'ne'; 1 instances of 'unke'; 2 instances of 'Ze'; 2 instances of 'I'; 2 instances of 'o'; 21 instances of 'iyoM'; 5 instances of 'isameM'; 4 instances of 'IM'; 1 instances of 'BI'; 2 instances of 's'; 1 instances of 'se'; 3 instances of 'us'; 2 instances of 'un'; 1 instances of 'ej'; 2 instances of 'isa'; 1 instances of 'MsoM'; 4 instances of 'eM'; 1 instances of 'jZa'; 2 instances of 'ismeM'; 1 instances of 'ere'; 1 instances of

'zvo'; 4 instances of 'iske'; 1 instances of 'ke'; 1 instances of 'usI'; 5 instances of 'is'; 1 instances of 'unmeM'; 2 instances of 'Ziyom'; 1 instances of 'in'; 1 instances of 'a';

- For tag **rsp**:
14 instances of 'kara'; 127 instances of "; 2 instances of 'M'; 6 instances of 'e'; 1 instances of 'I'; 1 instances of '1'; 1 instances of '0'; 1 instances of 'Z'; 1 instances of 'ne'; 1 instances of ';
- For tag **jjmod__intf**:
332 instances of "; 8 instances of 'I'; 1 instances of 'M'; 2 instances of 'Z'; 1 instances of 'Ze'; 1 instances of 'e';
- For tag **nmod__k2inv**:
75 instances of 'I'; 21 instances of "; 110 instances of 'e'; 8 instances of 'A'; 13 instances of 'uI'; 37 instances of 'ie'; 24 instances of 'ue'; 19 instances of 'i'; 4 instances of 'yA'; 10 instances of 'we'; 1 instances of 'vAyA'; 5 instances of 'Ze'; 1 instances of 'vAI'; 3 instances of 'iyA'; 2 instances of 'ZI'; 3 instances of 'ne'; 1 instances of 'Z'; 1 instances of 'ZwI'; 1 instances of 'uA'; 1 instances of 'ye'; 1 instances of 'iy'; 2 instances of 'wI';
- For tag **k2p**:
401 instances of "; 29 instances of 'M'; 1 instances of 'uneM'; 1 instances of 'ZoM'; 1 instances of 'iyom'; 9 instances of 'oM'; 7 instances of 'e'; 1 instances of 'un';
- For tag **ras-k2**:
19 instances of 'oM'; 17 instances of 'ne'; 111 instances of "; 13 instances of 'e'; 1 instances of 'inase'; 5 instances of 'iyom'; 1 instances of 'yom'; 1 instances of 'uske'; 1 instances of 'usse'; 1 instances of 'unke'; 1 instances of 'iske';
- For tag **pof__redup**:
164 instances of "; 6 instances of 'we'; 1 instances of 'Ze'; 20 instances of 'e'; 6 instances of 'I'; 1 instances of 'M'; 1 instances of 'Z'; 1 instances of 'i'; 1 instances of 'ZI';
- For tag **k5**:
130 instances of 'oM'; 833 instances of "; 3 instances of 'Ziyom'; 25 instances of 'M'; 27 instances of 'iyom'; 57 instances of 'e'; 2 instances of 'noM'; 1 instances of 'ZZ'; 18 instances of 'ne'; 25 instances of 'inmeM'; 4 instances of 'ismeM'; 1 instances of 'U'; 6 instances of 'isameM'; 9 instances of 'isse'; 7 instances of 'unse'; 1 instances of 'I'; 2 instances of 'se'; 6 instances of 'inameM'; 3 instances of 's'; 3 instances of 'inse'; 2 instances of 'Z'; 1 instances of 'use'; 1 instances of 'unmeM'; 2 instances of 'ja'; 3 instances of 'is'; 3 instances of 'usse'; 1 instances of 'n'; 1 instances of 'in'; 1 instances of 'IM'; 1 instances of 'eM'; 1 instances of 'yom'; 1 instances of 'unke'; 1 instances of 'Me'; 1 instances of 'isase';
- For tag **r6-k1**:
60 instances of 'oM'; 278 instances of "; 8 instances of 'Z'; 59 instances of 'ne'; 7 instances of 'uskA'; 18 instances of 'e'; 2 instances of 'is'; 8 instances of 'uskI'; 7 instances of 'Zne'; 1 instances of 'unke'; 16 instances of 'iskA'; 7 instances of 'iyom'; 5 instances of 'I'; 9 instances of 'unkI'; 6 instances of 'unkA'; 3 instances of 'iskI'; 2 instances of 'iske'; 3 instances of 'inkA'; 1 instances of 'vAne'; 2 instances of 'uske'; 1 instances of 'kA'; 2 instances of 'noM'; 1 instances of 'Ziyom'; 7 instances of 'M'; 1 instances of 'unoMne'; 2 instances of 'yom'; 1 instances of 'inaki'; 1 instances of 'isaki'; 1 instances of 'isakA'; 1 instances of 'ye';
- For tag **rd**:
226 instances of "; 14 instances of 'e'; 35 instances of 'oM'; 4 instances of 'I'; 1 instances of 'ArI'; 9 instances of 'iyom'; 1 instances of 'uskI'; 4 instances of 'unkI'; 1 instances of 'uske'; 1 instances of 'NULL'; 1 instances of 'unke'; 1 instances of 'ie'; 1 instances of 'yom';
- For tag **ras-k1**:
20 instances of 'unke'; 475 instances of "; 1 instances of 'ke'; 41 instances of 'iyom'; 109 instances of 'oM'; 4 instances of 'NULL'; 11 instances of 'e'; 6 instances of 'unse'; 1 instances of 'ere'; 4 instances of 'Are'; 5 instances of 'uske'; 3 instances of 'isse'; 9 instances of 'ne'; 3 instances of 'iske'; 4 instances of 'yom'; 2 instances of 'usse'; 1 instances of 'uom'; 1 instances of 'BI'; 1 instances of 'Ms'; 1 instances of 'oz'; 1 instances of 'se'; 1 instances of 'Ziyom'; 1 instances of 'is'; 1 instances of 'i';
- For tag **k7a**:
34 instances of 'iske'; 18 instances of 'e'; 430 instances of "; 170 instances of 'oM'; 3 instances of 'uske'; 12 instances of 'isake'; 2 instances of 'ne'; 13 instances of 'iyom'; 5 instances of 'isI'; 1 instances of 'i'; 9 instances of 'unke'; 2 instances of 'o'; 1 instances of 'Z'; 2 instances of 'Zne'; 1 instances of 'M';

- For tag **jjmod**:
788 instances of ”; 26 instances of ’ne’; 44 instances of ’e’; 7 instances of ’isse’; 43 instances of ’oM’; 13 instances of ’M’; 3 instances of ’ismeM’; 6 instances of ’Z’; 1 instances of ’unse’; 2 instances of ’;’; 1 instances of ’is’; 1 instances of ’Are’; 6 instances of ’unke’; 3 instances of ’usmeM’; 2 instances of ’iske’; 2 instances of ’0’; 4 instances of ’kara’; 1 instances of ’noM’; 1 instances of ’Zne’; 5 instances of ’iyoM’; 4 instances of ’usse’; 1 instances of ’ja’; 1 instances of ’NULL’; 2 instances of ’ie’; 1 instances of ’eM’; 1 instances of ’i’; 1 instances of ’Ms’; 2 instances of ’I’; 1 instances of ’uskI’;
- For tag **rh**:
6 instances of ’isI’; 683 instances of ”; 178 instances of ’ne’; 113 instances of ’isse’; 75 instances of ’oM’; 55 instances of ’e’; 6 instances of ’usse’; 11 instances of ’Z’; 61 instances of ’isase’; 9 instances of ’iyoM’; 1 instances of ’M’; 3 instances of ’isakI’; 8 instances of ’Zne’; 9 instances of ’ie’; 1 instances of ’yoMki’; 7 instances of ’yoM’; 1 instances of ’uske’; 13 instances of ’iske’; 3 instances of ’ZZ’; 15 instances of ’isake’; 1 instances of ’inakI’; 1 instances of ’inake’; 2 instances of ’inase’; 4 instances of ’iskI’; 1 instances of ’iyA’; 1 instances of ’unkI’; 2 instances of ’i’; 2 instances of ’I’; 1 instances of ’ZZiyoM’; 1 instances of ’eM’; 2 instances of ’kara’; 1 instances of ’Mne’; 1 instances of ’Ze’; 1 instances of ’is’; 1 instances of ’egI’;
- For tag **k3**:
374 instances of ”; 76 instances of ’oM’; 1 instances of ’ZZ’; 19 instances of ’e’; 5 instances of ’M’; 6 instances of ’isse’; 1 instances of ’usse’; 5 instances of ’isase’; 7 instances of ’iyoM’; 6 instances of ’isake’; 4 instances of ’unke’; 6 instances of ’iske’; 2 instances of ’Z’; 2 instances of ’NULL’; 2 instances of ’s’; 1 instances of ’;’; 2 instances of ’inake’; 2 instances of ’inase’; 1 instances of ’inIM’;
- For tag **nmod_k1inv**:
170 instances of ’I’; 304 instances of ’e’; 184 instances of ”; 101 instances of ’uI’; 18 instances of ’wI’; 104 instances of ’ue’; 17 instances of ’i’; 37 instances of ’A’; 75 instances of ’we’; 8 instances of ’Me’; 27 instances of ’ie’; 10 instances of ’gae’; 26 instances of ’Ze’; 34 instances of ’ne’; 1 instances of ’ZA’; 3 instances of ’yA’; 9 instances of ’ZI’; 3 instances of ’uA’; 3 instances of ’IM’; 3 instances of ’kara’; 3 instances of ’wA’; 7 instances of ’Zwe’; 9 instances of ’ZwI’; 1 instances of ’Zne’; 4 instances of ’iyA’; 1 instances of ’Z’; 3 instances of ’NULL’; 1 instances of ’gal’; 1 instances of ’j’; 1 instances of ’M’; 1 instances of ’ave’; 1 instances of ’iA’; 1 instances of ’MI’;
- For tag **k2g**:
34 instances of ”; 2 instances of ’oM’; 1 instances of ’ne’; 1 instances of ’noM’;
- For tag **ras-neg**:
45 instances of ”; 9 instances of ’e’; 1 instances of ’i’; 1 instances of ’iyoM’; 1 instances of ’M’; 2 instances of ’oM’; 7 instances of ’ne’; 1 instances of ’iske’; 1 instances of ’ue’;
- For tag **rbmod_relc**:
2 instances of ’M’; 5 instances of ’I’; 1 instances of ’wI’; 11 instances of ”; 2 instances of ’egA’; 2 instances of ’iyA’; 2 instances of ’uI’; 2 instances of ’e’; 1 instances of ’gA’;
- For tag **mod_wq**:
24 instances of ’is’; 35 instances of ”; 7 instances of ’i’; 4 instances of ’e’; 2 instances of ’I’;
- For tag **nmod_emph**:
43 instances of ”;
- For tag **k2u**:
34 instances of ”; 1 instances of ’Ze’; 2 instances of ’e’; 2 instances of ’oM’; 1 instances of ’iyoM’;
- For tag **k1u**:
65 instances of ”; 3 instances of ’e’; 3 instances of ’ne’; 12 instances of ’oM’; 3 instances of ’isse’; 2 instances of ’iyoM’; 1 instances of ’o’; 1 instances of ’usse’; 1 instances of ’ja’;
- For tag **rbmod**:
1 instances of ’I’; 31 instances of ”; 1 instances of ’is’;
- For tag **ras-r6**:
3 instances of ”;
- For tag **rtu**:
1 instances of ”;

- For tag **r6v**:
20 instances of ”; 2 instances of ’ne’; 1 instances of ’iskI’; 2 instances of ’oM’; 1 instances of ’inake’; 1 instances of ’ArA’;
- For tag **jk1**:
6 instances of ”; 1 instances of ’unse’; 2 instances of ’oM’; 1 instances of ’use’;
- For tag **rad**:
1 instances of ’ie’; 5 instances of ”;
- For tag **psp_cl**:
4 instances of ’e’; 2 instances of ”; 2 instances of ’I’;
- For tag **ras-k7**:
6 instances of ”; 4 instances of ’oM’; 1 instances of ’Z’;
- For tag **vmod_adv**:
1 instances of ”;
- For tag :
4 instances of ’ ’;
- For tag **k7u**:
8 instances of ”; 1 instances of ’e’; 1 instances of ’Z’; 1 instances of ’oM’;
- For tag **ras-k4**:
4 instances of ”; 1 instances of ’unke’; 1 instances of ’oM’;
- For tag **nmod_pofinv**:
2 instances of ’ul’; 5 instances of ’ie’; 3 instances of ’ue’; 2 instances of ’I’; 2 instances of ’e’; 1 instances of ’iyA’;
1 instances of ’ne’;
- For tag **k7pu**:
6 instances of ”;
- For tag **ras-k7p**:
5 instances of ”;
- For tag **pk1**:
6 instances of ”; 1 instances of ’oM’; 1 instances of ’unoMne’;
- For tag **k3u**:
1 instances of ”;
- For tag **k7tu**:
1 instances of ’e’; 1 instances of ’oM’; 1 instances of ”;
- For tag **ras_k7**:
2 instances of ”;
- For tag **jjmod_relc**:
1 instances of ’A’;
- For tag **mk1**:
1 instances of ’oM’;
- For tag **k4u**:
1 instances of ’oM’; 1 instances of ”;
- For tag **ras-r6-k2**:
1 instances of ’Z’; 1 instances of ”;
- For tag **ras-rt**:
1 instances of ”;

- For tag **ras-k4a**:
1 instances of ”;
- For tag **enm**:
1 instances of ”; 1 instances of 'e';
- For tag **undef**:
1 instances of ”;

3.2 Markers and Tag

For each marker, indicate the types of tags given to it and the number of cases for each tag. Example:

There are a total of 100 tokens with *-ne* marker. These tokens are marked with k_1, k_2 . k_1 marker is responsible for 32 of the cases and k_2 is responsible for the remaining 68.

3.3 N-gram of tags

Include statistics about the frequency of n-gram of tags. Take n in the range [2,4].

NOTE: You have to do the above 3 exercises for both UD and AnnCorra. This analysis has to be done on the training set given for both the types of tagging in Assignment 4.

4 Error Analysis for automatic tagging

In this section, describe the errors in the output of your model trained in Assignment 4 on the test data. Give possible solutions, if any, to mitigate these errors in the future.

5 Discussion

5.1 Comparison of UD and AnnCorra

5.2 Need for new tags

5.2.1 Intra-Chunk

propose any new intra-chunk tags you can think of with examples

5.2.2 Inter-Chunk

propose any new inter-chunk tags you can think of with examples.