CLG 452: Linguistic Data 2 **(Due: 27/04/20)**

Course Project Report

Instructor: Dr. Parameshwari *Name:* HARSHITA SHARMA, *Roll Number:* 20171099

1 Tags

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

1.1 UD

(i) nsubj and nsubj:pass - nominal subject

(1) **kabIra ne** kapade KZrIxe.
kabir.ERG clothes bought.PERF.PL
'Kabir bought clothes.'
Cue: nsubj is usually the subject of the sentence and holds the role of an agent. It can followed by markers 'ne', 'ko', 'se' or zero marker which can be followed by adjectives, verbs, nouns or pronouns.

(2) mAhI ke **bAla** kAle hEM.
mahi GEN hair.NOM black be.PRES.PL
'Mahi has black hair.'
Cue: nsubj need not be in the first position. And hence, to identify a nsubj relation one must look at the more agentive entity for the action.

(3) **arjuna xvArA** gIwa gAyA gayA.
arjun.ACC song.NOM sung PASSIVE
'Song was sung by Arjun.'
Cue: In case of a passive verb, the nominal subject is marked as nsubj:pass.

(ii) csubj: clausal subject

(1) **arjuna kA yaha socanA** sahI nahIM hE.
arjun.GEN this think.GERUND right NEG be.PRES
'This thinking of Arjun is not right.'
Cue: csubj is the clause that acts as the subject to the main verb.

(iii) obj: object

(1) arjuna ne **GadI** wodZI.
arjun ERG clock.ACC break.PERF.FEM
'Arjun broke the clock.'
Cue: The object in a sentence usually occurs in accusative case with a zero marker or a 'ko' marker. It plays the role of theme or patient.

(2) xAxI ne baccoM ko **kahAniyAz** sunAI.
grandmother ERG kids DAT stories tell.PERF.FEM
'Grandmother told stories to the children.'
Cue: In case of two objects, the direct object(theme) is marked as obj.

(iv) iobj: indirect object

(1) xAxI ne **baccoM ko** kahAniyAz sunAI.
grandmother ERG kids DAT stories tell.PERF.FEM
'Grandmother told stories to the children.'
Cue: As the name suggests, the indirect object is marked as iobj and occurs in dative case. It usually takes the role of recipient/beneficiary.

(v) obl: oblique nominal

(1) rAwa ko **AsamAna meM** wAre howe hEM.
night LOC_T sky LOC_P stars be.HAB.PL be.PRES.PL
'There are stars in the night sky.'
Cue: A noun, preposition or noun phrase that takes a role of an argument different than those taken by nsubj, obj and iobj are marked as obl. obl can be followed by markers- zero, 'mein', 'par', 'se', 'tak', 'ka/ke/ki' etc.

(2) **pedZ se** pawwe girawe hEM.
tree ABL leaves fall.HAB be.PRES.PL
'Leaves fall from the tree.'
Cue: obl includes entities with ablative, instrumental, locative cases too.

(3) unhoMne **xAla ke sAWa** cAvala KAe.
they.ERG curry with rice eat.PERF
'They ate rice with curry.'
Cue: obl includes associative cases too.

(vi) nmod: noun modifier

(1) **mAhI ke** bAla kAle hEM.
mahi GEN hair.NOM black be.PRES.PL
'Mahi has black hair.'
Cue: Usually, the entity in genetive case is marked as nmod.

(2) nEnA **apanI** bAwa spaRta rUpa se kahawI hE.
naina her point clearly say.HAB.FEM
'Naina tells her point clearly.'
Cue: Pronouns in genetive cases are also marked as nmod.

(vii) nummod: numeric modifier

(1) mAhI ke **wIna** Pala hEM.
mahi GEN three fruits be.PRES.PL
'Three fruits belong to Mahi.'
Cue: All numeric phrases are marked as nummod.

(viii) aux: auxiliary

(1) mAhI kabIra se milawI **hE**.
mahi.NOM kabir meet.FEM be.PRES.FEM
'Mahi meets Kabir.'
Cue: Helping verbs that denote the TAM and GNP in case of mostly habitual verbs are marked as aux.

(ix) aux:pass

(1) mAhI nAca rahI **hE**.
mahi.NOM dance CONT be.PRES
'Mahi is dancing.'
Cue: An extension of the helping verb that denote the TAM and GNP in case of mostly continuous TAM are marked as aux:pass.

(x) mark: marker

(1) mAhI ne kahA **ki** vaha kala Gara AegI.
mahi ERG said that she tomorrow home come.FUT.FEM
'Mahi said that she will come home tomorrow.'
Cue: Mainly, subordinating conjunctions are marked as mark.

(xi) cc: coordinating conjunction

(1) kabIra **Ora** nEnA KuSa We.
kabir and naina happy be.PERF
'Kabir and Naina were happy.'
Cue: Coordinating conjunctions are marked as ccof.

(xii) conj: conjunct

(1) kabIra Ora **nEnA** KuSa We.
kabir and naina happy be.PERF
'Kabir and Naina were happy.'
Cue: When a coordinating conjunction connects two or more elements, all elements in the relation except the first are marked as conj.

(xiii) root

(1) kabIra Ora nEnA KuSa **We**.
kabir and naina happy be.PERF
'Kabir and Naina were happy.'
Cue: The root of the sentence is marked root with a pseudo node as head.

(xiv) amod: adjectival modifier

(1) mAhI ke bAla **kAle** hEM.
mahi GEN hair.NOM black be.PRES.PL
'Mahi has black hair.'
Cue: Adjectives that modify a pronoun are marked as amod.

(2) vaha **suMxara** hE.
she beautiful be.PRES
'She is beautiful.'
Cue: Adjectives that modify a pronoun are also marked as amod.

(xv) advmod: adverbial modifier

(1) mEM **jalxI** uTawA hUz.
I.NOM early get up am
'I get up early.'
Cue: When an adverbial phrase modifies an entity or a clause/predicate it is annotated as advmod.

(2) arjuna kA yaha socanA sahI **nahIM** hE.
arjun.GEN this think.GERUND right NEG be.PRES
'This thinking of Arjun is not right.'
Cue: When an adverbial phrase modifies an entity(here, adjective 'sahI') or a clause/predicate it is annotated as advmod.

(xvi) cop: copula

(1) mAhI ke bAla kAle **hEM**.
mahi GEN hair.NOM black be.PRES.PL
'Mahi has black hair.'
Cue: 'hai' and its forms are the most common copula in Hindi. The verb whose objective is to equate two entities is marked as copula.

(xvii) dislocated

(1) **k**abIra (**raNabIra kapUra**) ko burA lagA.
kabir (ranbir kapoor) ACC hurt feel.PREF
'Kabir (Ranbir Kapoor) felt bad.'
Cue: When elements that are to occur together are separated by a set of punctuation/entities, the entity dislocated from its position is marked as dislocated.

(xviii) punct: punctuation

(1) mAhI ke bAla kAle hEM**.**
mahi GEN hair.NOM black be.PRES.PL
'Mahi has black hair.'
Cue: All punctuations are annotated as punct.

1.2 AnnCorra

(i) k1: *karta*, agent

(1) **arjuna** Gara gayA.
arjun.NOM home.ACC go.PERF
'Arjun went home.'
Cue: Usually karta is found in nominative case(0 marker) and the verb in active voice shows GNP agreement with karta.

(2) **kabIra ne** kapade KZrIxe.
kabir.ERG clothes bought.PERF.PL
'Kabir bought clothes.'
Cue: Karta has semantic role of a do-er. Whenever verb takes 'ya' TAM, karta take '-ne' marker.

(3) **mAhI ko** kiwAbeM paDZnI padI.
mahi.ACC books.NOM read.INF.FEM had.FEM
'Mahi had to read books.'
Cue: Whenever verb takes '-nA_padA' TAM, karta take 'ko' marker.

(4) **arjuna xvArA** gIwa gAyA gayA.
arjun.ACC song.NOM sung PASSIVE
'Song was sung by Arjun.'
Cue: Whenever verb takes '-nA_cAhiye' TAM, karta take 'ko' marker.

(5) **arjuna kA** kahanA hE kI mAhI suMxara hE.
arjun.GEN say is that mahi beautfiul be.PRES
'Arjun says that Mahi is beautiful.'
Cue: The karta occurs with a genitive postposition(kA) with verbs such as 'kaha', 'soca', 'mAna' etc. and the verb takes TAM '-nA'

(6) **arjuna kA yaha socanA** sahI nahIM hE.
arjun.GEN this think.GERUND right NEG be.PRES
'This thinking of Arjun is not right.'
Cue: Karta is the clause that acts as an agent to the main verb.

(7) arjuna ko **XanuRa** milA.
arjun DAT bow find.PERF
'Arjuna found a bow.'
Cue: The subject of an unaccusative verb acts as karta.

(8) **mEM** rojZ jalxI uTawA hUz.
I.NOM everyday early get up am
'I get up early every day.'
Cue: First and second person personal pronouns in nominative case act as karta.

(9) **mAhI** KUbasUrawa hE.
Mahi.NOM beautiful be.PRES
'Mahi is beautiful.'
Cue: Whenever stative verbs occur, the person or thing whose state is mentioned acts as the karta.

(ii) pk1: *prayojaka karta*, causer

(1) **nEnA ne** kabIra se kAma karavAyA.
naina.ERG kabir by work to do.CAUS
'Naina made Kabir work.'
Cue: pk1 is marked to a karta when the verb is cauative i.e. has '-A' or '-vA' suffix.

(2) **nEnA ne** munne ko Bojana KilAyA.
naina.ERG baby ACC food to eat.CAUS
'Naina made Kabir work.'
Cue: pk1 is marked to a karta when the verb is cauative i.e. has '-A' or '-vA' suffix.

(iii) jk1: *prayojya karta*, causee

(1) nEnA ne **kabIra se** kAma karavAyA.
naina ERG kabir by work to do.CAUS
'Naina made Kabir work.'
Cue: Usually, jk1 would have either a 'ko' marker or a 'se' marker.

(2) nEnA ne kabIra se **arjuna ko** nOkarI xilavAI.
naina ERG kabir by arjun.ACC job to give.CAUS.FEM
'Naina made Kabir to give a job to Arjun.'
Cue: The verb form and its semantics help determine the causee. Here, the ultimate causee is ***arjun****, while 'kabir' is just a mediator.*

(iv) mk1: *madhyastha karta*, mediator causer

(1) nEnA ne **kabIra xvArA** arjuna ko nOkarI xilavAI.
naina ERG kabir by arjun ACC job to give.CAUS.FEM
'Naina made Kabir to give a job to Arjun.'
Cue: The marker for a mk1 would either be 'xvArA' or 'se'. However, the the verb form and its semantics help best determine the mediator causer.

(2) nEnA ne **kabIra xvArA mahi se** arjuna ko nOkarI xilavAI
naina ERG kabir by mahi by arjun ACC job to give.CAUS.FEM
'Naina made Kabir to make Mahi give a job to Arjun.'
Cue: If more than one mk1 is present, then the first one would have 'xvArA' marker and the second one would have 'se' marker. However, the the verb form and its semantics help best determine the mediator causer.

(3) nEnA ne **kabIra xvArA** arjuna se kAma karavAyA.
naina ERG kabir by arjun by work to do.CAUS
'Naina made Kabir to make Arjun work.'
Cue: The second causer(a mediator) which is both a causee and a causer is marked as mk1.

(v) k1s: *vidheya karta - karta samanadhikarana*, noun complement of karta

(1) mAhI ke bAla **kAle** hEM.
mahi GEN hair.NOM black be.PRES.PL
'Mahi has black hair.'
Cue: k1s can only be there when a k1 is marked for a verb and has the same locus as karta.

(2) mAhI pahale iwanI **suMxara** nahIM WI.
mahi.NOM earlier so beautiful NEG be.PAST.FEM
'Mahi was not so beautiful earlier.'
Cue: k1s can only be there when a k1 is marked for a verb and has the same locus as karta.

(vi) k2: *karma*, object/patient

(1) arjuna ne **GadI** wodZI.
arjun ERG clock.ACC break.PERF.FEM
'Arjun broke the clock.'
Cue: Karma usually occurs either with a zero marker or a 'ko' marker and when both k1 and k2 occur with zero marker, k2 is the one that does not agree with the verb.

(2) kabIra ne **nEnA ko** pakadA.
kabir ERG naina.ACC catch.PERF
'Kabir caught Naina.'
Cue: Karma usually occurs either with a zero marker or a 'ko' marker and does not agree with the verb.

(3) mAhI **kabIra se** milawI hE.
mahi.NOM kabir meet.FEM be.PRES
'Mahi meets Kabir.'
Cue: Karma can also occur with a 'se' marker. In such occurrences, karma continues to take the semantics role of an object/patient.

(4) mAhI xvArA **sabjiyAz** kAtI gaIM.
mahi by vegetables.ACC cut.FEM PASSIVE.FEM.PL
'The vegetables were cut by Mahi.'
Cue: When the verb occurs with a passive TAM, then the noun which shows agreement with the verb is marked as k2.

(5) mAhI ne kahA **ki vaha kala Gara AegI**.
mahi ERG said that she tomorrow home come.FUT.FEM
'Mahi said that she will come home tomorrow.'
Cue: Finite clauses occur as sentential object hence act as k2.

(vii) k2p: goal, destination

(1) arjuna **Gara** gayA.
arjun.NOM home.ACC go.PERF
'Arjun went home.'
Cue: The object of motion verb is marked as k2p.

(2) nEnA rojZ **xaPZ_wara** AwI hE.
naina.NOM everyday office come.HAB.FEM be.PRES
'Naina comes to the office everyday.'
Cue: k2p doesn't agree with the verb unlike k2 under similar syntactic context. Here, ***xaPZ_wara*** *is a masculine noun but the verb,* ***AwI hE*** *is feminine.*

(viii) k2g: *secondary karma*

(1) vaha mAhI ko **lAdo** bulAwA hE.
he.NOM mahi ACC laado call.HAB be.PRES
'He calls Mahi as Laado.'
Cue: When a verb such as 'bulAnA' has two karma, the one which is dependent on the other k2 is marked as k2p.

(ix) k2s: *karma samanadhikarana*, object complement

(1) GaravAle nEnA ko **hoSiyAra** mAnawe hEM.
Family.NOM naina ACC intelligent believe.HAB be.PRES.PL
'The family believes Naina to be intelligent.'
Cue: k2s is a property that resides in k2 and acts as its complement. k2s can only be there if there is a k2 in a sentence.

(2) kabIra arjuna ko **kamajora** samaJawA hE.
kabir.NOM arjun ACC weak consider.HAB be.PRES
'Kabir considers Arjun to be weak.'
Cue: k2s is a property that resides in k2 and acts as its complement. k2s can only be there if there is a k2 in a sentence.

(x) k3: *karana*, instrument

(1) mAhI nEnA ko **rassI se** bAzXawI hE.
mahi.NOM naina ACC rope INST tie.FEM be.PRES
'Mahi ties Naina with a rope.'
Cue: karana karaka always takes a 'se' marker and takes the semantic role of instrument in achieving the action of the verb.

(xi) k4: *smapradana*, recipient

(1) mEMne **kabIra ko** upahAra xie.
I.ERG kabir DAT gifts give.PERF.PL
'I gave gifts to Kabir.'
Cue: k4 normally takes a 'ko' marker in Hindi.

(2) arjuna ne **mAhI se** kahA WA.
arjun ERG mahi to tell be.PAST
'Arjun told Mahi.'
Cue: In case of communication verbs, k4 takes 'ko' or 'se' as marker. The final destination of the action of the verb is marked as k4.

(3) xAxI ne **baccoM ko** kahAniyAz sunAI.
grandmother ERG kids DAT stories tell.PERF.FEM
'Grandmother told stories to the children.'
Cue: The final destination of the action of the verb is marked as k4.

(xii) k4a: *anubhava karta*, experiencer

(1) **arjuna ko** XanuRa milA.
arjun DAT bow find.PERF
'Arjuna found a bow.'
Cue: The argument of unaccusative verbs having a 'ko' marker is marked as k4a.

(2) **muJe** KAnA svAxiRta lagA.
I.DAT food delicious seem.PERF
'The food seemed delicious to me.'
*Cue: Verbs such as **laganA** 'to seem' and **xiKanA** 'to appear' take passive agents which would be marked as 'k4a'.*

(xiii) k5: *apadana*, source

(1) **pedZ se** pawwe girawe hEM.
tree ABL leaves fall.HAB be.PRES.PL
'Leaves fall from the tree.'
Cue: k5 always takes a 'se' marker. However, a better cue is to mark the entity representing the point of departure for a motion verb as k5.

(2) kabIra **nEnA se** nArAjZ hE.
kabir.NOM naina ABL angry be.PRES
'Kabir is angry with Naina.'
Cue: The entity which triggers emotions, in case of emotional verbs, is annotated as k5.

(xiv) k5prk: *prakruti apadana*, source material

(1) KIra **xUXa se** banawI hE.
rice-pudding milk from make.HAB.FEM be.PRES
'Kheer is made from milk.'
Cue: The participant in the action of verb which undergoes a change and also has a relation with the finished product is annotated as k5prk.

(xv) k7t: *kAlAdhikarana*, location in time

(1) **Aja** bAriSa hogI.
today.LOC_T rain be.FUT.FEM
'It will rain today.'
Cue: The participant denoting the time of action is marked as k7t. Usually k7t takes either zero or 'mein' or 'par' marker.

(2) **rAwa ko** AsamAna meM wAre howe hEM.
night LOC_T sky LOC_P stars be.HAB.PL be.PRES.PL
'There are stars in the sky at night.'
Cue: The participant denoting the time of action is marked as k7t. Usually k7t takes either zero or 'mein' or 'par' marker.

(3) **bacapana meM** saba mAsUma nahIM howe.
childhood LOC_T everyone innocent NEG be.HAB.PL
'Not everyone is innocent in childhood.'
Cue: The participant denoting the time of action is marked as k7t. Usually k7t takes either zero or 'mein' or 'par' marker.

(4) kabIra **vakZ_wa para** Gara AwA hE.
kabir.NOM time LOC_T home come.HAB be.PRES
'Kabir comes home on time.'
Cue: The participant denoting the time of action is marked as k7t. Usually k7t takes either zero or 'mein' or 'par' marker.

(5) mAhI **rAwa waka** vahIM bETI rahI.
mahi.NOM night LOC_T there.LOC_P sit.PERF.FEM stay.CONT
'Mahi sat there till night.'
Cue: The participant denoting the time of action is marked as k7t.

(xvi) k7p: *deshadhikarana*, location in space

(1) rAwa ko **AsamAna meM** wAre howe hEM.
night LOC_T sky LOC_P stars be.HAB.PL be.PRES.PL
'There are stars in the night sky.'
Cue: k7p refers to a location of karta or karma which is an actual physical place. Usually k7p takes either zero or 'mein' or 'par' marker.

(2) mAhI **vahIM** bETI rahI.
mahi.NOM there.LOC_P sit.PERF.FEM CONT
'Mahi remained sitting there.'
Cue: k7p refers to a location of karta or karma which is an actual physical place. Usually k7p takes either zero or 'mein' or 'par' marker.

(3) mAhI **mejZa para** bETI hE.
mahi.NOM table LOC_P sit.PERF.FEM be.PRES

'Mahi is sitting on the table.'
Cue: k7p refers to a location of karta or karma which is an actual physical place. Usually k7p takes either zero or 'mein' or 'par' marker.

(xvii) k7: *vishayadhikarana*, location elsewhere

(1) usake **ximAgZ meM** kuCa cala rahA WA.
he.GEN mind LOC_P something go.CONT CONT be.PAST
'There was something going on in his mind.'
*Cue: k7 refers to a metaphorical or abstract place not a physical place. Here, even though **ximAgZ** is a physical space but the thoughts aren't physically placed in the physical brain but in the mind.*

(2) kabIra Ora nEnA **carcA meM** hEM.
kabir and naina discussion LOC_P be.PRES.PL
'Kabir and Naina are a topic of discussion.'
Cue: k7 refers to a metaphorical or abstract place not a physical place.

(3) usakI **bAwoM para** XyAna mawa xo.
he.GEN words LOC_P attention NEG give.PRES
'Do not pay attention to his words.'
Cue: k7 refers to a metaphorical or abstract place not a physical place.

(xviii) k7a: according to

(1) **sUwroM ke anusAra** kabIra naI gAdI KZrIxegA.
sources GEN according kabir new vehicle buy.FUT
'According to sources, Kabir will buy a new vehicle.'
Cue: Noun chunks that take markers: 'ke$_m$ uwAbika', 'ke$_a$ nusAra' or 'ke$_w$ ahawa' are annotated as k7a.

(xix) k1u: *sAdrishya*, similarity or comparison of k1

(1) kabIra ko mAhI **nEnA kI wulanA mena** aXika nataKata lagI.
kabir DAT mahi naina GEN comparison in more naughty
appear.PERF.FEM
'Mahi appeared more naughty as compared to Naina to Kabir.'
Cue: The entity which k1 is being compared with is marked as k1u.

(2) kabIra ko mAhI **nEnA jEsI** nataKata lagI.
kabir DAT mahi naina like naughty appear.PERF.FEM
'Mahi appeared as naughty as Naina to Kabir.'
Cue: The entity to which k1 is being stated similar to is marked as k1u.

(xx) k2u: *sAdrishya*, similarity or comparison of k2

(1) kabIra mAhI ko **nEnA kI wulanA mena** aXika nataKata mAnawA hE.
kabir.NOM mahi ACC naina GEN comparison in more naughty
consider.HAB be.PRES
'Mahi appeared more naughty as compared to Naina to Kabir.'
Cue: The entity which k2 is being compared with is marked as k2u.

(2) kabIra mAhI ko **nEnA jEsI** nataKata samaJawA hE.
kabir.NOM mahi ACC naina like naughty consider.HAB be.PRES
'Mahi appeared as naughty as Naina to Kabir.'
Cue: The entity to which k2 is being stated similar to is marked as k2u.

(xxi) r6: *shashthi*, genitive or possessive

(1) **mAhI ke** bAla kAle hEM.
mahi GEN hair.NOM black be.PRES.PL
'Mahi has black hair.'
Cue: The genitive or possessive relation which holds between two nouns is marked as r6. r6 takes 'ka'/'ke'/'ki' markers.

(xxii) rd: *prati*, direction

(1) **maMxira kI waraPZ** eka kuMA hE.
temple GEN direction one well be.PRES
'There is a well towards the temple.'
Cue: An element having markers such as 'kI_ora' or 'ke_prati' is marked as rd.

(2) **CAwroM ke prawi** aXyApikA kA hqxaya komala ho gayA.
students GEN direction teacher GEN heart soft be.PERF go.PERF
'The teacher's heart turned soft towards the students.'
Cue: An element having markers such as 'kI_ora' or 'ke_prati' is marked as rd.

(xxiii) rh: *hetu*, reason

(1) **nEnA ke kAraNa** saba vahAz pahuzce.
naina GEN because everyone there reach.PERF.PL
'Everyone reached there because of Naina.'
Cue: An element indicating the reason of action is marked as rh. rh usually takes 'ke_kAraNa', 'kI_vajaha_se', 'se' as markers or when a conjunct 'kyoMki' is present.

(xxiv) rt: *tadarthya*, purpose

(1) mAhI ne **arjuna ke lie** kZmIjZ KZrIxI.
mahi ERG arjun for shirt buy.PERF
'Mahi bought shirt for Arjun.'
Cue: The entity that refers to the purpose of action is annotated as rt. Usually 'ke_liye' marker indicates a rh relation.

(xxv) ras-k1: *upapada$_s$ahakArakatwa*, associative

(1) nEnA **kabIra ke sAWa** GUmane ke lie manAlI gaI.
naina.NOM kabir with travel for manali.LOCP go.PERF
'Naina went to Manali with Kabir for travelling.'
Cue: ras-k takes markers 'ke$_s$AWa', 'ke sAWa sAWa', and 'kI waraha'. The entity that is being associated with someone decides the k*.*

(xxvi) ras-k2: *upapada$_s$ahakArakatwa*, associative with k2

(1) unhoMne **xAla ke sAWa** cAvala KAe.
they.ERG curry with rice eat.PERF
'They ate rice with curry.'
Cue: ras-k takes markers 'ke$_s$AWa', 'ke sAWa sAWa', and 'kI waraha'. The entity that is being associated with someone decides the k*.*

(xxvii) ras-neg: Negation in Associatives

(1) para vaha **nEnA ke binA** calA gayA.
but.ERG he.NOM naina without go.PERF
'But he left without Naina.'
Cue: ras-neg takes markers 'ke binA' conveying the sense of doing an action without the ras-neg marked entity.

(xxviii) pof: part of units(conjunct verbs)

(1) mAhI ke GaravAloM ne arjuna ko **svIkAra karA**.
mahi GEN family ERG arjun ACC accept do.PERF
'Mahi's family accepted Arjun.'
Cue: All conjunct verbs are annotated with pof. Hence, in a pof relation: a noun/adjective is attached to a verb to give the verb a new meaning.

(xxix) ccof: co-ordination and sub-ordination

(1) kabIra **Ora** nEnA KuSa We.
kabir and naina happy be.PERF
'Kabir and Naina were happy.'
Cue: Coordinating conjunctions are marked as ccof.

(2) mAhI ne kahA **ki** vaha kala Gara AegI.
mahi ERG said that she tomorrow home come.FUT.FEM
'Mahi said that she will come home tomorrow.'
Cue: Subordinating conjunctions are also marked as ccof.

(xxx) rsym: symbol

(1) mAhI ke GaravAloM ne arjuna ko svIkAra karA**.**
mahi GEN family ERG arjun ACC accept do.PERF
'Mahi's family accepted Arjun.'
Cue: In hindi: A vertical line is a special symbol for 'full-stop' marked as rsym. Apart from this most if the English symbols are present in Hindi are are annotated as rsym.

2 Linguistic Challenges with Annotation

(a)Differential Object marking In Differential Object marking, objects are deferentially marked with different case markers, depending on which, they show agreement with a verb. For example:

(1) mAhI ne *kiwAba* **paDI**
mahi ERG book read.PERF
'Mahi read a book.'

(2) mAhI ne *kiwAba ko* **paDA**
mahi ERG book read.PERF
'Mahi read a book.'

As seen in the above mentioned examples, when the object takes zero marker it shows GNP agreement with the verb whereas when it 'ko' it changes it behaviour with the verb and the agreement is not seen anymore. If we were to add **definiteness** to the first example by adding a definite pronoun 'usa', we find that the object 'kiwAba' takes a case marker 'ko':

(3) mAhI ne usa *kiwAba ko* **paDA**
mahi ERG that book read.PERF
'Mahi read that book.'

Similarly, **alienability** also plays a role in differential object marking. For example:

(4) wuma rAma kI **kiwAbeM** le Ao
you ram GEN books bring come
'Bring Ram's books.'

(5) wuma rAma kI **bahanoM ko** le Ao
you ram GEN sisters bring come
'Bring Ram's sisters.'

Animacy also determines whether the object will take marker or not.

(6) sIwA ne mele meM **cUdiyAz** *xeKIM*
sita ERG carnival LOCP bangles see.PERF.PL
'Sita saw bangles in the carnival'

(7) sIwA ne mele meM **rAma ko** *xeKA*
sita ERG carnival LOCP ram see.PERF
'Sita saw Ram in the carnival.'

Challenges: As seen in the examples above, because of DOM, the syntactic cues do not always apply hence, cause problems with annotation.

(b) Non-Nominative Subjects Unlike many languages, in Hindi, subjects are not just identified in nominative case, but they also occur in ergative case, dative case etc. and hence the agreement is determined by factors other than the GNP of the syntactic subject.

(1) **arjuna ko** XanuRa milA.
arjun DAT bow find.PERF
'Arjuna found a bow.'

(2) **sIwA ne** mele meM **rAma ko** xeKA
sita ERG carnival LOCP ram see.PERF
'Sita saw Ram in the carnival.'

Challenges: As seen in the examples above, because of the presence of non-nominative subjects in the language, not only the syntactic cues and positions fail in determining the subject but the entity showing agreement with the verb also changes and hence, cause problems with annotation. On top of that the phenomenon of **split ergativity** that takes the 'ne' marker only in some tenses, increase the confusion.

(c) Complex Predicates Complex predicates or complex verbs occur in the form of **conjunct verbs** i.e. nominal+verb and **compound verbs** i.e. verb+verb. For example:

(1) usane kahAnI **SurU kI**.
he story start do.PERF
'He started the story.'

(2) sIwA ne KAnA **KA liyA**
sita ERG carnival LOCP ram see.PERF
'Sita saw Ram in the carnival.'

Challenges: Example 1 has a conjunct verb and example 2 has a compound verb. In case of conjunct verbs, while annotating an issue arises - whether the nominal is part of the complex verb or is it an overt argument of the verb. Similarly, while annotating compound verbs, the problem of whether the second verb is an aspectual/auxiliary/modal or is the verb complex a combination of verb and intensifier. When coupled with verb phrase ellipsis they increase the problem with annotation by a fair share.

(d) Non-finite clauses: Conditional, Concessive, Relative, participial clauses Non-finite clause refer to those clauses which are based on an infinitive verb or a participle and has no tense.

(1) agara mEM **jA sakawA** wo mEM jAwA
if I go can.PERF then I go.PAST
'If I could go I would go'

(2) hAlAzki jo mEM nahIM **cAhawA**, vahI howA hE
although what I NEG want.HAB, that happen.HAB be
'Although whatever I don't want, that happens.'

(3) jo ladZkA **nAca rahA hE**, vaha merA BAI hE
the boy dance CONT be.PRES, he my brother be.PRES
'The boy who is dancing is my brother.'

(4) usane **BAgawe hue** Sera ko xeKA
he running lion ACC see.PERF
'He saw a running lion.'

Challenges: As seen in the examples, the concept of finiteness in terms of verbs is missing altogether. These clauses portray a different picture but are different in spirit, like 'BAgawe hue' is acting as an adjective to lion but are simply infinite clauses in picture. Hence, pose a challenge while annotating.

(e) Ambiguity (Coordination, Attachment)

1. Coordination Ambiguity: Whether all entities bound in a conjunct relation get modified or the entity with minimal attachment. For example: javAna ladZkA Ora ladZkI has two possible meanings:
   (a) **javAna ladZkA** Ora ladZkI - Here 'javAna' applies only to 'ladZkA'
   (b) **javAna ladZkA** Ora **ladZkI** - - Here 'javAna' applies to both 'ladZkA' and 'ladZkI'
2. Attachment Ambiguity occurs when a part of sentence can attach to two different heads giving two different meanings. For example, rAma ne BAgawe hue Sera ke bacce ko xeKA can have a number of meanings:
   (a) Ram saw a **running lion cub**
   (b) **Ram** saw a lion cub while **running**
   (c) Ram saw a **running lion**'s cub

Challenges: Since a number of possible meanings arise, a number of annotation possibilities also arise.

(f) Ellipsis When a part of sentence is omitted from the sentence which can be retrieved through context and meaning. For example:

1. Ram gaya, Mohan nahi
2. Ram khana chahta hai, Mohan nahi chahta
3. Ram apne kutte ko ghumane lata hai, Mohan bhi

Challenges: In Ellipsis, especially pseudo-ellipsis, where TAM is retained but the verb is ellided(example 2), annotation becomes difficult.

(g) Non-projectivity Non-projectivity occurs when a word or a phrase is separated from another word or phrase that it modifies in a way that a direct connection between the two is not possible without incurring crossing lines in the tree structure.

(h) Particles Particles do not change with GNP and are not included in the grammatical classes like noun, verb etc. Some examples:

1. bhi: main bhi raja hoon(inclusion), aur bhi pyara(intensity)
2. hi: tu hi hai(singleton); tu hi bhagwaan hai(focus)

3. tak: raat tak(throughout), raat ko soyi tak nahi(not even)
4. bhar: pal bhar(entire)

Challenges: Since particles are widely used, one particle can exhibit more than one property through which ambiguities arise and cause problems in annotation.

PTO

3 Tag Statistics

3.1 Markers and Tag

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

3.1.1. AnnCorra: For each *vibhakti* the types of tags and number of cases are listed below:

Table 1: Markers and Tags for AnnCorra data

marker	tag	#cases
vAloM ko	k4	8
	k2	3
	k1	2
	ccof	2
vAloM ke	r6	2
	vmod	1
vAloM kI	r6	5
	ccof	1
	nmod	1
vAloM kA	r6	1
	k1	1
vAloM	ccof	1
vAle para	k7	1
vAle ne	k1	2
vAle ko	k2	1
vAle ke lie	rt	2
vAle ke	vmod	3
	r6	1
	k7	1
vAle	nmod	352
	lwg__vaux	35
	nmod__k1inv	15
	k1	12
	ccof	10
	k1s	2
	nmod__relc	1
	r6	1
	pof__cn	1
	nmod__k2inv	1
	nmod__pofinv	1
	jjmod	1
	k7t	1
	ras-neg	1
ke	k7u	1
	ras-k4	1
	ras-k7p	1
	nmod__k1inv	1
	ras-r6	1
	fragof	1

marker	tag	#cases
se lekara	rsp	16
	nmod	4
	vmod	3
	ccof	2
se pare	k1s	1
	nmod	1
se alaga	vmod	1
sahiwa	ccof	26
	vmod	14
	ras-k1	9
	ras-k2	7
	nmod	2
	rsym	1
	ras-k7	1
	ras-k7p	1
	adv	1
samewa	vmod	50
	ccof	43
	ras-k1	14
	ras-k2	5
	nmod	3
	ras-pof	1
	ras-k7	1
	ras-r6	1
	rsym	1
samAna	k1u	1
	k2s	1
samaya	k7t	1
vAloM se	k5	1
	k4	1
vAloM meM	k7	8
vAloM para	k7	2
vAloM ne	k1	2
jEsI	ccof	13
	nmod	9
	k2u	4
	rsym	3
	k7t	1
	k7u	1

Table 2: Markers and Tags for AnnCorra data cotinued

marker	tag	#cases
se	k5	1037
	adv	445
	k2	443
	rh	401
	lwg__psp	303
	k7t	298
	k3	297
	k4	275
	ras-k1	235
	jjmod	228
	ccof	166
	nmod	164
	rsp	120
	k7p	79
	k7	62
	vmod	61
	rt	46
	ras-k2	41
	lwg__vaux	30
	rsym	28
	k1	11
	k1u	10
	rd	7
	k1s	7
	k2p	7
	k2s	6
	r6	5
	k2g	5
	lwg__rp	4
	pof__redup	4
	k2u	4
	jk1	3
	k4a	3
	k7a	3
	k7u	2
	pof	2
	sent-adv	2
	ras-r6-k2	2
	ras-rt	1
	ras-k4a	1
	nmod__adj	1
	ras-pof	1
sI	k1s	1
jarie	k3	9
	rh	1

marker	tag	#cases
vAloM ke viruxXa	vmod	1
vAloM ke lie	rt	4
vAloM ke bAre	k7	1
vAloM ke KilAPZ	vmod	1
meM bawOra	k7p	3
	k7t	1
muwAbika	k7a	13
	k7	1
maxxenajZra	k7a	1
maxxenajara	k7a	4
	jjmod	1
bAvajUxa	vmod	10
	k7a	4
bAre meM	k7	105
bAbawa	k7	5
	nmod	1
	k2	1
banAma	nmod	1
baxale	vmod	3
bawOra	k2	2
	lwg__rp	2
	nmod	2
	main	1
	vmod	1
	ccof	1
	nmod__relc	1
bakOla	¡/s¿	10
POYra	pof__cn	1
waka ke lie	rt	8
	k7t	5
waka ke	r6	6
	k7t	1
waka kI	r6	7
	r6-k2	4
	r6-k1	3
	rsym	1
waka kA	r6	11
	r6-k2	2
jAnabUJakara	k1	1
ke rUpa meM	k7	2
	k7p	1
	k1u	1
ke muwAbikZ	k7a	10
	k7	1

Table 3: Markers and Tags for AnnCorra data cotinued

marker	tag	#cases
meM	k7	4180
	k7p	2827
	k7t	670
	ccof	263
	lwg__psp	182
	jjmod	137
	nmod	131
	lwg__rp	70
	rsym	42
	adv	22
	k2p	18
	k2	14
	lwg__vaux	8
	pof	7
	k1	7
	sent-adv	5
	rt	4
	rd	4
	vmod	3
	k1s	2
	pof__redup	1
	pof__cn	1
	r6	1
	k5	1
	k2s	1
	nmod__adj	1
	k2u	1
	k1u	1
me	k7	1
vAlI	nmod	204
	lwg__vaux	16
	nmod__k1inv	12
	ccof	11
	nmod__k2inv	1
	r6	1
	jjmod	1
vAlA	nmod	36
	k1s	7
	ccof	4
	lwg__vaux	3
	k1	2
	nmod__k1inv	2
	rsym	1
	k2s	1
vAloM xvArA	k1	1

marker	tag	#cases
ke KilAPZ	jjmod	1
	vmod	1
ke sivA	ccof	1
ke silasile meM	k7	1
ke sAWa	ccof	1
ke samAnAMwara	nmod	1
ke samAna	k1u	5
	k7p	1
	k1s	1
	nmod	1
	k7pu	1
ke samaya meM	k7t	1
ke samaya para	k7t	1
ke samaya	k7t	1
ke saMbaMXa meM	k7	5
ke viroXasvarUpa	rh	1
ke viruxXa	vmod	4
ke viparIwa	k1s	1
ke vAswe	rt	3
ke liye	rt	1
ke muwAbika	k7a	266
	ccof	1
	rsym	1
ke mukAbale	k1u	2
	k2u	1
	k7	1
ke mArPawa	adv	1
ke maxxenajZra	k7a	5
	rsym	1
	ccof	1
ke maxxenajara	k7a	25
	vmod	4
	k7	3
	lwg__vaux	1
	ccof	1
	rsym	1
ke maxxejanara	k7a	1
ke BI	k1	1
ke bAvajUxa	vmod	62
	lwg__vaux	5
	k7a	3
ke bAbawa	k2	1
	vmod	1
ke bAxa	k7t	2

Table 4: Markers and Tags for AnnCorra data cotinued

marker	tag	#cases
ko	k2	2769
	k4	1439
	k7t	1009
	k1	367
	ccof	288
	k4a	243
	rt	93
	rsym	42
	jjmod	15
	nmod	13
	k7p	12
	lwg__vaux	11
	vmod	5
	jk1	4
	lwg__rp	4
	k2g	4
	k2s	3
	r6	3
	nmod__emph	2
	k7	2
	k5	1
	k2p	1
	rh	1
	r6-k1	1
	k1s	1
	lwg__psp	1
	nmod__k2inv	1
	rs	1
ke jZrie	k3	6
ke lie	rt	1565
	ccof	75
	nmod	13
	lwg__vaux	13
	jjmod	12
	rh	10
	r6	9
	rsym	8
	k7t	5
	k2	4
	k4	3
	k1	2
	lwg__rp	2
	vmod	1
	k1s	1
calawe	rh	12
	vmod	6

marker	tag	#cases
ke barAbara	vmod	2
	k1s	2
	rsym	2
ke baxale meM	k7	2
ke baxale	pof__cn	2
	vmod	2
	nmod	1
	rsym	1
	k7	1
ke bajAya	vmod	10
	ras-neg	8
	ccof	1
ke bajAe	ras-neg	4
	vmod	1
	k7	1
ke PalasvarUpa	ccof	1
ke prawi	rd	45
	ccof	4
	nmod	2
	k2	2
	k7	1
ke pariNAmasvarUpa	rh	2
ke nAwe	rh	2
	vmod	2
	ccof	1
	k7	1
ke xvArA	k1	5
	k3	3
	ccof	2
ke wOra para	k7	5
	vmod	2
ke wahawa	k7a	125
	ccof	6
	rsym	6
ke jEsA	k1u	1
ke jariye	k3	18
	lwg__rp	2
	ccof	1
ke jarie	k3	78
	ccof	6
	rsym	1
	rh	1
	k7	1
ke calawe	rh	37
	ccof	2
	lwg__vaux	1

Table 5: Markers and Tags for AnnCorra data cotinued

marker	tag	#cases
lie	rt	141
	k4a	1
	r6	1
	vmod	1
lAyaka	vmod	2
	k1s	1
	k1	1
meM se	k5	37
	k7	14
	nmod	4
	ccof	3
	k7p	1
prawi	rd	2
	k3	1
	rt	1
para se	ccof	1
	k5	1
	k7	1
para	k7	2364
	k7p	615
	ccof	102
	lwg__vaux	63
	k7t	54
	jjmod	54
	nmod	33
	lwg__psp	32
	k2	26
	adv	16
	vmod	15
	rsym	14
	k2p	11
	pof__redup	7
	rh	6
	k1	3
	rt	2
	nmod__adj	2
	k4	1
	main	1
	k5	1
ne bawOra	k1	1
wahawa	k7a	31
	k7	2
	vmod	1
wale	k7	2
	k7p	1
waka se	k5	1
waka meM	k7	2
ke KilAPZ	vmod	8
	k2	1
ke KilAPa	vmod	89
	ccof	6
	jjmod	2
	rsym	1
ke kAraNa	rh	10
	ccof	2
	lwg__vaux	1
ke evaja meM	vmod	1
ke ulata	vmod	2
ke alAvA	vmod	68
	ccof	19
	rsym	1
ke anusAra	k7a	126
ke anurUpa	k7a	5
	vmod	1
ke anukUla	vmod	2
ke aXIna	k7a	2
	k7	1
	vmod	1
ke awirik2̆00dwa	k7a	1
	vmod	1
ke awirikwa	vmod	3
ke awaMrgawa	k7	1
ke aMrwagawa	k7a	1
ke aMwargawa	k7a	7
kI vajaha se	rh	2
	ccof	1
kI bAbawa	lwg__vaux	1
	nmod	1
	rh	1
kI baxOlawa	k7	1
kI bajAya	ras-neg	3
kI bajAe	ccof	1
kI wulanA meM	k7	1
OYPa	pof__cn	5
OYna	pof__cn	7
ulata	sent-adv	1
ina	pof__cn	1
anusAra	k7a	3
awirikwa	k7p	1

Table 6: Markers and Tags for AnnCorra data cotinued

marker	tag	#cases
alAvA	vmod	112
	sent-adv	7
	k7a	2
kA	r6	2007
	r6-k2	1044
	k1	300
	r6-k1	177
	ccof	141
	rsym	37
	lwg__vaux	29
	nmod	11
	r6v	7
	lwg__psp	7
	ras-k2	4
	k2	4
	lwg__rp	3
	pof	3
	pof__cn	3
	k1s	2
	jjmod	1
	lwg__vaux_cont	1
	nmod__adj	1
	rh	1
	vmod	1
	k7t	1
	ras-k1	1
kI waraha	k1u	5
	adv	3
	nmod	1
	k7pu	1
	k4u	1
	rtu	1
ke bAre meM	k7	192
	ccof	21
	lwg__vaux	5
	k2	3
	nmod	3
	lwg__psp	2
	r6	1
	k7p	1
KilAPa	vmod	15
	jjmod	1
	k7a	1
ko bawOra	k2	1

marker	tag	#cases
kI	r6	4784
	r6-k2	915
	ccof	312
	r6-k1	188
	rd	182
	lwg__vaux	106
	rsym	70
	rh	53
	nmod	39
	k1u	23
	k7	21
	jjmod	21
	k2u	12
	k1	11
	pof__cn	11
	lwg__psp	10
	lwg__rp	10
	k2	7
	r6v	6
	ras-k2	5
	k5	5
	rt	4
	k7t	4
	pof__redup	4
	nmod__adj	3
	k7u	3
	k3	3
	k7tu	3
	adv	3
	vmod	3
	nmod__k1inv	2
	lwg__vaux_cont	2
	k1s	2
	k4u	1
	k7p	1
	ras-k4	1
	ras-k1	1
	ras-neg	1
kAraNa	rh	1
jEsA	nmod	6
	k2u	3
	k1u	2
	adv	1
	k7pu	1

Table 7: Markers and Tags for AnnCorra data cotinued

marker	tag	#cases
ke	r6	6312
	k7t	986
	k7	557
	ccof	531
	ras-k1	320
	rh	190
	k7p	182
	vmod	178
	k1	160
	rsym	124
	r6-k2	88
	lwg__vaux	81
	ras-k2	80
	nmod	55
	r6-k1	41
	k7a	39
	jjmod	31
	ras-neg	29
	k2	24
	pof__cn	20
	lwg__rp	19
	k5	15
	k1u	14
	k3	11
	lwg__psp	11
	adv	11
	r6v	8
	rt	6
	ras-k7	6
	k2p	5
	rs	4
	k4	4
	k2s	3
	sent-adv	3
	ras-pof	3
	rd	3
	rsp	2
	nmod__adj	2
	k1s	2
	pof	2
	pof__redup	2
	k4a	1
	ras__k7	1
	k3u	1

marker	tag	#cases
waka	k7t	440
	k7	79
	k7p	33
	k2p	20
	sent-adv	17
	lwg__psp	14
	ccof	11
	vmod	9
	k1s	8
	nmod	6
	k2	6
	k1	5
	pof	3
	lwg__vaux	3
	jjmod	2
	rs	2
	adv	2
	pof__redup	1
	k2s	1
	k4	1
	k5	1
	r6	1
jEse	ccof	44
	nmod	23
	k1u	7
	rsym	5
	k2u	4
	k7u	3
	lwg__rp	2
	k7pu	1
	lwg__vaux	1
xvArA	k1	294
	jjmod	40
	k3	29
	ccof	19
	nmod	11
	rsym	8
	k2	3
	r6-k1	1
	mk1	1
	jk1	1
	k7p	1
jariye	k3	1
bAre meM	k7	105

Table 8: Markers and Tags for AnnCorra data cotinued

marker	tag	#cases
ne	pk1	3
	r6	2
	k2	1
	r6-k1	1

marker	tag	#cases
ne	k1	4916
	ccof	144
	rsym	48
	lwg__rp	4

3.1.2. UD: For each *case marker/adposition* the types of tags and number of cases are listed below:

Table 9: Markers and Tags for UD data

marker	tag	#cases
hewu	obl	1
havAle se	obl	1
svarUpa	obl	1
se lekara	nmod	17
	conj	1
se bawOra	obl	1
se pare	nmod	1
	obj	1
se alaga	nmod	1
se	obl	2700
	obj	426
	nmod	425
	iobj	264
	case	188
	conj	148
	advcl	138
	advmod	57
	punct	25
	nummod	20
	aux:pass	14
	aux	13
	acl	11
	nsubj	11
	xcomp	6
	root	6
	mark	6
	compound	6
	nsubj:pass	4
	amod	4
	dep	3
	acl:relcl	1
sI	xcomp	1
sahiwa	obl	28
	conj	18
	nmod	7
	punct	1

marker	tag	#cases
sarIKI	nmod	2
sarIKA	obl	1
samewa	obl	62
	conj	26
	nmod	6
	punct	1
samAna	root	1
	acl	1
samaya	obl	13
	advcl	2
saMbaXI	amod	1
saMbaMXI	amod	6
	nmod	2
	aux:pass	2
viroXa meM	obl	1
viparIwa	obl	1
vAloM se	iobj	1
	obl	1
vAloM meM	obl	7
vAloM para	obl	1
vAloM ne	nsubj	2
vAloM xvArA	nsubj:pass	1
vAloM ko	iobj	9
	obj	3
	conj	2
	nsubj	2
vAloM ke viruxXa	advcl	2
vAloM ke lie	advcl	3
	obl	1
vAle ke	nmod	1
	obl	1
sarIKe	nmod	1
vAloM kI	nmod	5
	amod	1
	conj	1

Table 10: Markers and Tags for UD data continued

marker	tag	#cases
vAloM kA	nsubj	1
	nmod	1
vAloM	obj	1
vAle samaya meM	obl	1
vAle para	obl	1
vAle ne	nsubj	1
vAle ko	obj	1
vAle ke lie	advcl	2
vAle ke bAre	obl	1
vAle ke KilAPZ	advcl	1
vAle ke KilAPa	advcl	2
vAle ke	nmod	1
	obl	1
vAle	amod	263
	nmod	76
	aux:pass	25
	acl	16
	nsubj	12
	conj	7
	aux	5
	obl	2
	xcomp	1
	compound	1
	acl:relcl	1
	root	1
vAlI	amod	145
	nmod	51
	acl	13
	aux:pass	13
	conj	10
	aux	2
	obl	1
	nummod	1
vAlA	amod	23
	nmod	10
	obj	5
	acl	4
	conj	3
	aux:pass	3
	nsubj	2
	punct	1
vajaha se	obl	7
	nmod	1
vakwa	obl	4
baxale	obl	2

marker	tag	#cases
lie	obl	119
	nmod	14
	nsubj	1
lAyaka	obj	1
	nsubj	1
	advcl	1
	advmod	1
	amod	1
yogya	obj	1
meM se	obl	35
	nmod	16
	conj	3
	obj	1
	root	1
meM bawOra	obl	4
meM	obl	6651
	nmod	707
	conj	228
	root	68
	dep	64
	punct	38
	obj	38
	case	15
	compound	8
	advmod	7
	nsubj	7
	acl:relcl	6
	aux	6
	advcl	5
	amod	4
	aux:pass	2
	nsubj:pass	1
	acl	1
me	obl	1
muwAbika	obl	7
	nmod	5
mAXyama se	obl	1
maxxenajZra	obl	1
maxxenajara	obl	4
bAvajUxa	obl	7
	nmod	2
	root	1
bAre meM	obl	96
	nmod	3
banAma	nmod	1

Table 11: Markers and Tags for UD data cotinued

marker	tag	#cases
bAbawa	obl	4
	nmod	2
	obj	1
bawOra	dep	2
	advcl	1
	conj	1
	obj	1
	nummod	1
	amod	1
	acl:relcl	1
	mark	1
bakOla	punct	10
PalasvarUpa	obl	1
prawi	obl	4
pariNAmasvarUpa	obl	2
para se	conj	1
	obl	1
para	obl	2738
	nmod	187
	conj	95
	aux:pass	49
	root	34
	obj	28
	advcl	17
	punct	14
	aux	11
	acl:relcl	8
	compound	7
	nsubj	3
	nummod	2
	advmod	1
	amod	1
	iobj	1
ne bawOra	nsubj	1
ne	nsubj	4683
	conj	137
	punct	48
	nsubj:pass	10
	dep	3
	obl	3
	obj	1
waka se	obl	1
waka meM	obl	2
waka ke lie	obl	13
jZrie	obl	1

marker	tag	#cases
xvArA	nsubj:pass	195
	nsubj	85
	obl	64
	conj	18
	nmod	15
	punct	7
	obj	4
wahawa	obl	26
	nmod	2
	root	1
wale	obl	3
waraha	dep	1
waka ke	nmod	6
waka kI	nmod	13
	punct	1
waka kA	nmod	13
waka	obl	528
	nmod	40
	case	11
	obj	7
	conj	6
	root	6
	xcomp	6
	compound	4
	nsubj	3
	aux:pass	3
	advcl	2
	nummod	2
	mark	2
	nsubj:pass	1
	acl	1
	iobj	1
	amod	1
jEse	conj	35
	nmod	23
	obl	10
	punct	3
	amod	1
	aux	1
	root	1
jEsI	nmod	9
	conj	8
	obl	8
	punct	3
	obj	2

Table 12: Markers and Tags for UD data cotinued

marker	tag	#cases
jEsA	nmod	5
	obl	5
	obj	1
	root	1
jariye	obl	1
jarie	obl	11
	nmod	1
jagaha	nmod	1
calawe	obl	14
	nmod	1
KilAPa	obl	31
	nmod	3
KilAPZ	obl	2
ko bawOra	obj	1
ko	obj	2614
	iobj	1338
	obl	1016
	nsubj	541
	conj	263
	advcl	80
	punct	40
	nsubj:pass	36
	nmod	35
	aux:pass	8
	acl	7
	dep	4
	aux	3
	amod	1
	xcomp	1
	root	1
ke hisAba se	obl	2
	nmod	1
ke havAle se	obl	11
ke sOjanya se	nsubj:pass	1
ke sivA	conj	1
ke silasile meM	obl	1
ke samAnAMwara	nmod	1
ke samAna	obl	5
	nmod	3
	obj	2
ke samaya se	punct	1
ke samaya meM	obl	2
ke samaya para	obl	2
ke mukAbale	obl	1
ke mArPawa	obl	1

marker	tag	#cases
ke samaya	obl	35
	nmod	7
	conj	2
	aux:pass	1
ke samakRa	obl	1
ke saMbaMXa meM	obl	25
	nmod	2
	aux:pass	2
ke saMbaMXa	obl	1
ke viroXasvarUpa	obl	1
ke viroXa meM	obl	2
	aux:pass	1
ke virUxXa	conj	1
ke viruxXa	obl	5
	conj	1
ke viparIwa	xcomp	1
ke vAswe	advcl	3
ke vakwa	obl	7
	nmod	2
ke lihAja se	conj	1
	obl	1
ke liye	advcl	1
ke lie	obl	762
	advcl	637
	nmod	113
	conj	68
	punct	8
	aux	8
	obj	6
	aux:pass	5
	iobj	3
	root	3
	amod	2
	nsubj	2
	dep	2
	advmod	1
ke rUpa meM	obl	97
	nmod	5
	conj	4
	punct	2
ke rupa meM	obl	1
	punct	1
ke mukAbale meM	obl	1
ke mAXyama se	obl	2
ke maxxejanara	obl	1

Table 13: Markers and Tags for UD data cotinued

marker	tag	#cases
ke muwAbikZ	obl	11
	nmod	1
ke muwAbika	obl	224
	nmod	34
	punct	1
	conj	1
ke mukAbale	obl	18
	nmod	6
	nummod	1
ke maxxenajZra	obl	5
	conj	1
	punct	1
ke maxxenajara	obl	27
	nmod	3
	conj	1
	punct	1
	advcl	1
	aux:pass	1
ke bUwe para	obl	1
ke bUwe	conj	1
ke bAvajUxa	obl	29
	advcl	28
	aux:pass	5
	nmod	5
ke bAre meM	obl	175
	conj	20
	nmod	16
	aux:pass	4
	obj	3
	mark	2
	aux	1
ke bAbawa	advcl	1
	obj	1
	obl	1
ke bahAne	obl	2
	obj	1
ke bala para	obl	2
ke calawe	obl	32
	advcl	3
	conj	2
	nmod	1
	aux:pass	1
ke ulata	obl	2
ke KilAPZ	obl	7
ke anurUpa	obl	9

marker	tag	#cases
ke barAbara	obl	2
	xcomp	1
	root	1
	punct	1
ke baxale meM	obl	2
ke baxale	obl	2
	punct	1
	advcl	1
	nmod	1
ke bajAya	obl	10
	advcl	6
	conj	1
ke bajAe	obl	5
	advcl	1
ke PalasvarUpa	conj	1
ke prawi	obl	32
	nmod	11
	conj	3
	obj	2
ke paScAwa	obl	1
ke pariNAmasvarUpa	obl	5
ke nAwe	advcl	5
	obl	1
	conj	1
ke xvArA	nsubj:pass	5
	obl	2
	conj	2
ke wOra para	obl	25
	xcomp	1
ke wahawa	obl	112
	nmod	6
	conj	6
	punct	5
ke wawvAvaXAna meM	obl	1
ke jEsA	obl	1
ke jZrie	obl	6
ke jariye	obl	17
	dep	1
	conj	1
	nmod	1
ke jarie	obl	79
	conj	5
	punct	1
ke AXAra para	obl	6
ke anukUla	obl	2

Table 14: Markers and Tags for UD data cotinued

marker	tag	#cases
ke alAvA	obl	46
	conj	14
	nmod	10
	advcl	7
	punct	1
ke anusAra	obl	114
	nmod	31
	root	3
ke evaja meM	obl	3
	advcl	1
ke KilAPZ	obl	15
	nmod	4
	obj	2
	conj	1
ke KilAPa	obl	156
	conj	19
	nmod	15
	obj	2
	advcl	1
	punct	1
ke kAraNa	obl	115
	advcl	67
	conj	16
	aux	10
	nmod	6
	aux:pass	3
ke aXIna	obl	4
	nmod	1
ke awirik2̆00dwa	nmod	2
ke awirikwa	nmod	1
	obl	1
ke aMrwagawa	obl	1
ke aMwargawa	obl	7
kI baxOlawa	obl	1
kI bajAya	obl	4
	aux:pass	1
kI bajAe	conj	1
kI wulanA meM	obl	15
	nmod	1
kI warja para	obl	1
kI KAwira	obl	2
	advcl	1
kI apekRA	nmod	2
	obl	1
kAraNa	obl	14

marker	tag	#cases
ke	nmod	6248
	obl	1852
	conj	462
	nsubj	146
	punct	115
	aux:pass	57
	obj	25
	advcl	19
	dep	15
	aux	13
	root	12
	case	10
	nsubj:pass	8
	amod	6
	acl:relcl	4
	compound	4
	iobj	4
	acl	3
	advmod	3
	nummod	1
kI hEsiyawa se	obl	1
kI vajaha se	obl	37
	advcl	16
	conj	4
	aux	3
	aux:pass	1
	punct	1
kI vajaha	obl	1
	nmod	1
kI BAMwi	obl	3
	nmod	1
kI bAbawa	obl	3
	nmod	1
	aux:pass	1
kI waraha	obl	38
	punct	3
	advmod	2
	conj	2
	amod	1
	obj	1
	nmod	1
kI waraPa se	obl	1
kI jagaha	obl	6
	nmod	1
ulata	obl	2

Table 15: Markers and Tags for UD data cotinued

marker	tag	#cases
kI	nmod	5588
	conj	285
	obl	200
	aux:pass	78
	punct	65
	aux	21
	case	10
	nsubj	10
	amod	8
	obj	7
	dep	6
	root	5
	compound	4
	nsubj:pass	2
	nummod	2
	xcomp	1
	acl	1
	mark	1
	advcl	1
alAvA	obl	95
	nmod	18

marker	tag	#cases
kA	nmod	2981
	nsubj	282
	conj	133
	punct	36
	obl	17
	aux	14
	aux:pass	14
	case	5
	obj	5
	dep	3
	amod	2
	compound	2
	root	1
	nsubj:pass	1
	acl	1
	xcomp	1
	advmod	1
anusAra	obl	3
awirikwa	nmod	1
vAloM ke	nmod	2

3.2 Tag and Markers

For each tag, indicate the markers used to identify that tag and the number of tokens identified by each marker.

3.2.1. AnnCorra: For each *tag* the types of markers and number of cases are listed below:

Table 16: Tags and Markers for AnnCorra data

marker	tag	#cases
sent-adv	null	309
	waka	17
	meM	17
	kA	13
	ke	12
	kI	10
	se	7
	alAvA	7
	para	7
	ko	5
	samaya	1
	ulata	1
	ke lie	1
	ke KilAPZ	1
rtu	kI waraha	1

marker	tag	#cases
rsp	se	120
	null	17
	se lekara	16
	ke	2
rs	null	1476
	ke	4
	waka	2
	ko	1
ras-r6	samewa	1
	ke	1
	null	1
ras__k7	ke	1
	null	1
undef	null	1
vmod__adv	null	1

Table 17: Tags and Markers for AnnCorra data continued

marker	tag	#cases
rsym	null	20245
	ke	46
	kI	30
	meM	24
	kA	19
	ko	19
	ne	18
	se	14
	para	6
	ke lie	6
	xvArA	3
	ke wahawa	3
	ke maxxenajZra	1
	ke alAvA	1
	ke jarie	1
	sahiwa	1
	jEsI	1
	jEse	1
	waka kI	1
	ke maxxenajara	1
rh	null	521
	se	401
	ke	196
	kI	60
	ke calawe	37
	calawe	12
	ke lie	10
	ke kAraNa	10
	para	7
	meM	6
	ko	4
	kA	3
	ke pariNAmasvarUpa	2
	ke nAwe	2
	jarie	1
	waka	1
	vajaha se	1
	ke viroXasvarUpa	1
	ne	1
	kAraNa	1
	ke jarie	1
	kI bAbawa	1
	ke bAvajUxa	1
rbmod_relc	null	28
ras-k4a	se	1

marker	tag	#cases
r6	ke	6313
	kI	4785
	null	3153
	kA	2008
	waka kA	11
	ke lie	9
	se	8
	waka kI	7
	ko	7
	waka ke	6
	meM	6
	vAloM kI	5
	ne	5
	para	2
	vAloM ke	2
	ke alAvA	2
	vAlI	1
	ke xvArA	1
	vAle	1
	vAloM kA	1
	vAle ke	1
	lie	1
	jEsA	1
	waka	1
r6-k2	kA	1044
	kI	915
	null	339
	ke	88
	waka kI	4
	waka kA	2
	meM	2
	ne	1
	vAle	1
r6-k1	kI	188
	kA	177
	null	108
	ke	41
	waka kI	3
	xvArA	1
	ko	1
	ne	1
ras-rt	se	1
ras-r6-k2	se	2
rbmod	null	33
rad	null	6

Table 18: Tags and Markers for AnnCorra data continued

marker	tag	#cases
rt	ke lie	1565
	null	195
	lie	141
	ko	96
	se	48
	ke	7
	kI	5
	meM	5
	para	4
	ke vAswe	3
	kA	2
	ne	2
	hewu	1
	ke liye	1
	prawi	1
	bAre meM	1
rd	kI	182
	null	45
	ke prawi	45
	se	8
	meM	8
	ke	3
	kA	2
	ko	2
	prawi	2
	bAre meM	1
	ke lie	1
ras-pof	ke	3
	samewa	1
	null	1
	se	1
ras-neg	ke	29
	null	22
	ke bajAya	8
	ke bajAe	4
	kI bajAya	3
	vAle	1
	kI	1
ras-k7p	null	3
	sahiwa	1
	ke	1
ras-k7	ke	6
	null	3
	samewa	1
	sahiwa	1

marker	tag	#cases
ras-k2	ke	80
	se	42
	null	17
	sahiwa	7
	ko	6
	kA	5
	kI	5
	samewa	5
	meM	1
	ke jarie	1
	vAlI	1
	ne	1
ras-k1	ke	323
	se	241
	null	86
	samewa	14
	sahiwa	9
	meM	7
	ko	5
	kI	5
	para	4
	kA	4
	ne	3
	ke lie	2
	waka	1
	se bawOra	1
	vAlI	1
r6v	ke	8
	kA	7
	kI	6
	null	6
psp__cl	null	8
pof__redup	null	182
	para	7
	se	4
	kI	4
	ke	2
	meM	1
	waka	1
pof__cn	null	19067
pk1	null	5
	ne	3
nmod__relc	null	1023
	vAle	1
	bawOra	1

Table 19: Tags and Markers for AnnCorra data continued

marker	tag	#cases
main	null	14087
	bawOra	1
	para	1
lwg__vaux_cont	null	5710
	ke	4
	kI	2
	kA	1
lwg__vaux	null	13079
	kI	106
	ke	82
	para	63
	vAle	35
	se	31
	kA	29
	vAlI	16
	ko	15
	ke lie	13
	meM	9
	ke bAvajUxa	5
	vAlA	3
	waka	3
	ke kAraNa	1
	kI bAbawa	1
	ke maxxenajara	1
	jEse	1
	ke calawe	1
lwg__rp	null	4015
	meM	71
	ke	42
	kI	14
	se	9
	ko	4
	ne	4
	kA	3
	ke lie	3
	ke jariye	2
	bawOra	2
	jEse	2
lwg__neg	null	1945
k7u	jEse	3
	kI	3
	se	2
	jEsI	1
	ke	1
	null	1

marker	tag	#cases
k7t	null	2373
	ko	1070
	ke	1040
	meM	732
	waka	450
	se	350
	para	74
	kI	55
	ne	45
	kA	44
	ke lie	11
	vAlA	3
	vAle	3
	ke bAvajUxa	2
	ke bAxa	2
	vAloM kI	1
	samaya	1
	jEsI	1
	meM bawOra	1
	vAlI	1
	vAle ne	1
	ke xvArA	1
	ke muwAbika	1
	waka ke	1
	bAre meM	1
	ke anusAra	1
	kI waraha	1
	ke samaya	1
k7p	meM	2829
	null	840
	para	616
	ke	186
	se	90
	waka	34
	ko	12
	meM bawOra	3
	kA	3
	kI	2
	meM se	1
	wale	1
	ke samAna	1
	ne	1
	xvArA	1
	awirikwa	1
mod__wq	null	72

Table 20: Tags and Markers for AnnCorra data continued

marker	tag	#cases
nmod	null	4045
	vAle	352
	vAlI	204
	se	171
	meM	135
	ke	56
	kI	41
	para	36
	vAlA	36
	jEse	23
	ke lie	14
	ko	14
	kA	12
	xvArA	12
	jEsI	9
	jEsA	6
	waka	6
	se lekara	4
	meM se	4
	samewa	3
	sarIKI	2
	sahiwa	2
	ke prawi	2
	bawOra	2
	ke baxale	1
	kI waraha	1
	ke samAnAMwara	1
	kI bAbawa	1
	vAloM kI	1
	bAbawa	1
	ne	1
	banAma	1
	bAre meM	1
	se pare	1
	ke samAna	1
k4	ko	1439
	null	402
	se	275
	vAloM ko	8
	ke	4
	ke lie	3
	para	1
	waka	1
	vAloM se	1
	ne	1

marker	tag	#cases
k7	meM	4216
	para	2375
	null	977
	ke	607
	bAre meM	105
	se	88
	waka	87
	kI	81
	ko	43
	ne	21
	kA	14
	meM se	14
	ke lie	12
	vAloM meM	8
	bAbawa	5
	ke muwAbika	3
	ke maxxenajara	3
	wahawa	2
	waka meM	2
	vAle	2
	xvArA	2
	vAloM para	2
	wale	2
	vAlI	1
	muwAbika	1
	ke prawi	1
	ke aXIna	1
	ke wahawa	1
	ke nAwe	1
	ke bajAe	1
	ke awaMrgawa	1
	kI baxOlawa	1
	vAle para	1
	ke muwAbikZ	1
	me	1
	samewa	1
	vAle ke	1
	ke baxale	1
	ke jarie	1
	ke muwAbikZ	1
	ke mukAbale	1
	para se	1
k2g	null	29
	se	5
	ko	4

Table 21: Tags and Markers for AnnCorra data continued

marker	tag	#cases
k2	null	11028
	ko	2769
	se	443
	para	28
	ke	26
	meM	14
	kI	7
	waka	6
	ne	5
	kA	4
	ke lie	4
	vAloM ko	3
	xvArA	3
	bawOra	2
	ke prawi	2
	ke bajAe	1
	ke bAbawa	1
	vAle ko	1
	ke KilAPZ	1
	bAbawa	1
	ko bawOra	1
k1u	kI	23
	ke	14
	null	13
	se	10
	jEse	7
	jEsI	5
	ke samAna	5
	kI waraha	5
	jEsA	2
	ke mukAbale	2
	samAna	1
	ke jEsA	1
	sarIKA	1
	para	1
	meM	1
nmod__emph	null	41
	ko	2
nmod__adj	null	24392
	ke	6
	kI	3
	para	2
	se	1
	meM	1
mod	null	2264

marker	tag	#cases
k1s	null	2549
	se	10
	waka	8
	vAlA	7
	ke	3
	kA	2
	vAle	2
	para	2
	meM	2
	ke barAbara	2
	kI	2
	se pare	1
	ke samAna	1
	lAyaka	1
	ke lie	1
	ko	1
	sI	1
	ke viparIwa	1
k1	null	13215
	ne	4917
	ko	369
	kA	301
	xvArA	294
	ke	160
	vAle	12
	se	12
	meM	11
	kI	11
	waka	5
	ke xvArA	5
	para	4
	vAloM ne	2
	vAloM ko	2
	ke lie	2
	vAle ne	2
	vAlA	2
	ne bawOra	1
	alAvA	1
	ke BI	1
	vAloM kA	1
	vAloM xvArA	1
	lAyaka	1
fragof	null	280
	ke	1
enm	null	2

Table 22: Tags and Markers for AnnCorra data continued

marker	tag	#cases
ccof	null	16746
	ke	538
	kI	316
	ko	290
	meM	265
	se	169
	ne	146
	kA	141
	para	102
	ke lie	76
	jEse	44
	samewa	43
	sahiwa	26
	xvArA	19
	ke alAvA	19
	jEsI	13
	waka	12
	vAlI	12
	vAle	10
	ke jarie	6
	ke wahawa	6
	ke KilAPa	6
	ke prawi	4
	vAlA	4
	meM se	3
	se lekara	2
	ke calawe	2
	ke xvArA	2
	vAloM ko	2
	ke kAraNa	2
	vAloM	1
	vAloM kI	1
	ke bajAya	1
	ke sivA	1
	ke muwAbika	1
	ke sAWa	1
	ke nAwe	1
	bawOra	1
	ke jariye	1
	ke maxxenajara	1
	para se	1
	ke maxxenajZra	1
	kI bajAe	1
	ke PalasvarUpa	1
k3u	ke	1
k7tu	kI	3

marker	tag	#cases
k2s	null	395
	se	6
	ko	3
	ke	3
	vAlA	1
	waka	1
	meM	1
	samAna	1
k2p	null	388
	waka	20
	meM	18
	para	11
	se	7
	ke	5
	ko	1
jjmod__relc	null	1
jjmod__intf	null	345
jjmod	null	418
	se	229
	meM	139
	para	54
	xvArA	40
	ke	31
	kI	22
	ko	16
	ke lie	12
	waka	2
	ke KilAPa	2
	ne	2
	kA	1
	maxxenajara	1
	KilAPa	1
	vAlI	1
	vAle	1
	ke KilAPZ	1
adv	null	1454
	se	444
	meM	23
	para	16
	ke	12
	kI waraha	3
	kI	3
	waka	2
	ke mArPawa	1
	jEsA	1
	sahiwa	1

Table 23: Tags and Markers for AnnCorra data continued

marker	tag	#cases
k5	se	1037
	null	81
	meM se	37
	ke	15
	kI	5
	ko	2
	vAloM se	1
	meM	1
	waka	1
	para se	1
	waka se	1
	para	1
k4u	kI	1
	kI waraha	1
k4a	ko	243
	null	162
	se	3
	ke	1
	lie	1
k3	se	297
	ke jarie	78
	null	61
	xvArA	29
	ke jariye	18
	ke	11
	jarie	9
	ke jZrie	6
	kI	3
	ke xvArA	3
	para	2
	meM	1
	prawi	1
	ko	1
	jZrie	1
	jariye	1
k2u	kI	12
	null	11
	jEse	4
	se	4
	jEsI	4
	jEsA	3
	ke mukAbale	1
	meM	1
ras-k4	null	3
	kI	2
	ke	1

marker	tag	#cases
nmod__pofinv	null	15
	vAle	1
nmod__k2inv	null	340
	vAlI	1
	vAle	1
	ko	1
nmod__k1inv	null	1139
	vAle	15
	vAlI	12
	kI	2
	vAlA	2
	ke	1
k7pu	kI waraha	1
	jEse	1
	null	1
	sarIKe	1
	ke samAna	1
	jEsA	1
k7a	ke muwAbika	266
	ke anusAra	126
	ke wahawa	125
	ke	39
	wahawa	31
	ke maxxenajara	25
	null	25
	muwAbika	13
	ke muwAbikZ	10
	ke aMwargawa	7
	ke maxxenajZra	5
	ke anurUpa	5
	bAvajUxa	4
	maxxenajara	4
	ke bAvajUxa	3
	anusAra	3
	se	3
	ke aXIna	2
	alAvA	2
	ke maxxejanara	1
	maxxenajZra	1
	ke aMrwagawa	1
	ke awirikž00dwa	1
	KilAPa	1
jk1	ko	4
	se	3
	null	2
	xvArA	1

Table 24: Tags and Markers for AnnCorra data continued

marker	tag	#cases
pof	null	11567
	meM	7
	waka	3
	kA	3
	ke	2
	se	2
	kI	1
vmod	null	2794
	ke	181
	alAvA	113
	ke KilAPa	89
	ke alAvA	68
	se	66
	ke bAvajUxa	62
	samewa	50
	para	19
	KilAPa	15
	sahiwa	14
	bAvajUxa	10
	ke bajAya	10
	waka	9
	ne	9
	ko	8
	ke KilAPZ	8
	kI	7
	calawe	6
	meM	6

marker	tag	#cases
vmod	kA	5
	ke maxxenajara	4
	ke viruxXa	4
	ke awirikwa	3
	vAle ke	3
	se lekara	3
	baxale	3
	ke barAbara	2
	ke nAwe	2
	lAyaka	2
	ke ulata	2
	ke baxale	2
	KilAPZ	2
	ke anukUla	2
	ke bajAe	1
	bawOra	1
	ke lie	1
	ke anurUpa	1
	ke bAbawa	1
	ke aXIna	1
	wahawa	1
	xvArA	1
	vAloM ke	1
	se alaga	1
	ke awirik200dwa	1
	ke KilAPZ	1
	ke muwAbika	1
	lie	1

3.2.2. UD: For each *tag* the types of markers and number of cases are listed below:

Table 25: Tags and Markers for UD data

marker	tag	#cases
acl:relcl	null	941
	para	8
	meM	6
	ke	4
	se	1
	vAle	1
	bawOra	1
case	null	3436
	se	188
	meM	15
	waka	11
	ke	10
	kI	10
	kA	5

marker	tag	#cases
dislocated	null	37
det	null	5832
dep	null	3822
	meM	64
	ke	15
	kI	6
	ko	4
	ne	3
	se	3
	kA	3
	bawOra	2
	ke lie	2
	ke jariye	1
	waraha	1

Table 26: Tags and Markers for UD data continued

marker	tag	#cases
compound	null	28897
	meM	8
	para	7
	se	6
	kI	4
	ke	4
	waka	4
	kA	2
	vAle	1
acl	null	2487
	vAle	16
	vAlI	13
	se	11
	ko	7
	vAlA	4
	ke	3
	waka	1
	kI	1
	meM	1
	kA	1
	samAna	1
aux:pass	null	8568
	kI	79
	ke	58
	para	49
	vAle	25
	se	15
	kA	14
	vAlI	13
	ko	10
	ke lie	5
	ke bAvajUxa	5
	ke kAraNa	3
	vAlA	3
	waka	3
	meM	3
	saMbaMXI	2
	xvArA	2
	ke calawe	1
	ke samaya	1
	kI bajAya	1
	ke maxxenajara	1
	kI bAbawa	1
cop	null	2736

marker	tag	#cases
xcomp	null	603
	se	6
	waka	6
	sI	1
	kI	1
	vAlI	1
	ke viparIwa	1
	ke barAbara	1
	vAle	1
	ko	1
	kA	1
root	null	13157
	meM	68
	para	34
	ke	12
	waka	6
	se	6
	kI	5
	ke anusAra	3
	ke lie	3
	kA	1
	samAna	1
	wahawa	1
	bAvajUxa	1
	ke barAbara	1
	ko	1
	jEsA	1
	meM se	1
	jEse	1
	vAle	1
aux	null	9119
	kI	21
	se	14
	kA	14
	ke	14
	para	11
	ke kAraNa	10
	ke lie	8
	meM	6
	vAle	5
	ko	5
	vAlI	2
	jEse	1
vocative	null	6

Table 27: Tags and Markers for UD data continued

marker	tag	#cases
punct	null	18293
	ke	116
	kI	66
	ne	48
	ko	41
	meM	39
	kA	36
	se	25
	para	14
	bakOla	10
	ke lie	8
	xvArA	7
	ke wahawa	5
	jEse	3
	kI waraha	3
	jEsI	3
	ke muwAbika	1
	ke baxale	1
	ke maxxenajara	1
	samewa	1
	ke bAvajUxa	1
	ke KilAPa	1
	ke alAvA	1
	sahiwa	1
	vAlA	1
	ke barAbara	1
	ke jarie	1
	waka kI	1
	ke maxxenajZra	1
mark	null	5915
	se	6
	waka	2
	kI	1
	bawOra	1
iobj	ko	1338
	null	374
	se	264
	vAloM ko	9
	ke	4
	ke lie	3
	vAloM se	1
	waka	1
	para	1

marker	tag	#cases
nsubj	null	11208
	ne	4683
	ko	541
	kA	282
	ke	146
	xvArA	85
	vAle	12
	se	11
	kI	10
	meM	7
	waka	3
	para	3
	vAloM ko	2
	ke lie	2
	vAloM ne	2
	vAlA	2
	lAyaka	1
	lie	1
	vAloM kA	1
	ne bawOra	1
	vAle ne	1
nummod	null	4444
	se	20
	para	2
	kI	2
	waka	2
	bawOra	1
	vAlI	1
	ke	1
	ke mukAbale	1
nsubj:pass	null	214
	xvArA	195
	ko	36
	ne	10
	ke	8
	ke xvArA	5
	se	4
	kI	3
	waka	1
	meM	1
	kA	1
	vAloM xvArA	1

Table 28: Tags and Markers for UD data continued

marker	tag	#cases
advcl	null	2561
	ke lie	637
	se	141
	ko	82
	ke kAraNa	67
	ke bAvajUxa	28
	ke	23
	para	20
	meM	10
	ke alAvA	7
	ke bajAya	6
	ke nAwe	5
	kA	4
	ke calawe	3
	ke vAswe	3
	kI	3
	ne	3
	waka	2
	samaya	2
	lAyaka	1
	bawOra	1
	bAre meM	1
	ke bAbawa	1
	ke liye	1
	ke KilAPa	1
	ke viruxXa	1
	ke maxxenajara	1
	kI KAwira	1
	ke baxale	1
	ke bajAe	1
obj	null	10412
	ko	2614
	se	426
	meM	38
	para	29
	ke	26
	waka	7
	kI	7
	ke lie	6
	vAlA	5
	kA	5
	xvArA	4
	vAloM ko	3
	ke samAna	2
	ke prawi	2
	ne	2

marker	tag	#cases
obj	ke KilAPa	2
	ke KilAPZ	2
	jEsI	2
	yogya	1
	lAyaka	1
	vAloM	1
	bawOra	1
	meM se	1
	ko bawOra	1
	jEsA	1
	bAbawa	1
	kI waraha	1
	ke bAbawa	1
	vAle ko	1
	ke bahAne	1
	se pare	1
amod	null	10927
	vAle	263
	vAlI	145
	vAlA	23
	kI	8
	saMbaMXI	6
	ke	6
	meM	4
	se	4
	ke lie	2
	kA	2
	vAloM kI	1
	ko	1
	kI waraha	1
	bawOra	1
	jEse	1
	lAyaka	1
	saMbaXI	1
	para	1
	waka	1
conj	null	3827
	ke	472
	kI	293
	ko	264
	meM	228
	se	150
	ne	139
	kA	136
	para	95
	ke lie	68

Table 29: Tags and Markers for UD data continued

marker	tag	#cases
conj	null	3827
	jEse	35
	samewa	26
	ke KilAPa	19
	xvArA	18
	sahiwa	18
	ke kAraNa	16
	ke alAvA	14
	vAlI	11
	jEsI	8
	waka	7
	vAle	7
	ke wahawa	6
	ke jarie	5
	vAlA	3
	ke prawi	3
	meM se	3
	ke samaya	2
	vAloM ko	2
	kI waraha	2
	ke calawe	2
	ke xvArA	2
	kI bajAe	1
	ke nAwe	1
	ke bajAya	1
	para se	1
	ke viruxXa	1
	vAloM kI	1
	ke maxxenajara	1
	bawOra	1
	ke KilAPZ	1
	ke virUxXa	1
	ke maxxenajZra	1
	ke bUwe	1
	ke sivA	1
	ke bajAe	1
	ke PalasvarUpa	1
	ke jariye	1
	se lekara	1
	ke muwAbika	1
nmod	null	10407
	ke	6252
	kI	5594
	kA	2982
	meM	712
	se	426
	para	190
	ke lie	113
	vAle	76
	vAlI	51
	waka	40
	ko	37
	ke muwAbika	34
	ke anusAra	31
	jEse	23
	alAvA	18
	se lekara	17
	meM se	16
	xvArA	16
	ke KilAPa	15
	lie	14
	waka kA	13
	waka kI	13
	ke prawi	11
	ke alAvA	10
	vAlA	10
	jEsI	9
nmod	sahiwa	7
	ke samaya	7
	ke kAraNa	6
	ke wahawa	6
	waka ke	6
	samewa	6
	ke mukAbale	6
	jEsA	5
	muwAbika	5
	vAloM kI	5
	ke bAvajUxa	5
	bAre meM	4
	ke KilAPZ	4
	KilAPa	3
	ke samAna	3
	ke maxxenajara	3
	bAvajUxa	2
	ke awirikž00dwa	2
	saMbaMXI	2
	ke vakwa	2
	kI apekRA	2
	vAloM ke	2
	bAbawa	2

Table 30: Tags and Markers for UD data continued

marker	tag	#cases
obl	meM	6714
	null	6337
	para	2756
	se	2739
	ke	1934
	ko	1079
	ke lie	775
	waka	534
	kI	276
	ke muwAbika	230
	ke KilAPa	156
	lie	120
	ke kAraNa	116
	ke anusAra	114
	ke wahawa	114
	bAre meM	96
	alAvA	95
	ke jarie	80
	xvArA	66
	samewa	63
	ke alAvA	46
	kI waraha	38
	meM se	37
	ke samaya	36
	ne	35
	ke prawi	33
	ke calawe	32
	KilAPa	31
	ke bAvajUxa	30
	kA	29
	sahiwa	28
	ke maxxenajara	27
	wahawa	26
	ke mukAbale	18
	ke jariye	17
	ke KilAPZ	15
	kAraNa	14
	calawe	14
	samaya	13
	jarie	11
	ke muwAbikZ	11
	jEse	10
	ke bajAya	10
	ke anurUpa	9
	jEsI	8
	bAvajUxa	7

marker	tag	#cases
obl	muwAbika	7
	ke aMwargawa	7
	ke vakwa	7
	ke KilAPZ	7
	vAloM meM	7
	vajaha se	7
	kI jagaha	6
	ke jZrie	6
	ke samAna	5
	ke pariNAmasvarUpa	5
	ke bajAe	5
	ke viruxXa	5
	ke maxxenajZra	5
	prawi	5
	jEsA	5
	ke aXIna	4
	meM bawOra	4
	maxxenajara	4
	bAbawa	4
	vakwa	4
	kI bajAya	4
	vAle	4
	kI BAMwi	4
	wale	3
	kI bAbawa	3
	anusAra	3
	vAlI	3
	ke barAbara	3
	waka meM	2
	vAlA	2
	ke baxale	2
	ke ulata	2
	baxale	2
	KilAPZ	2
	ulata	2
	ke xvArA	2
	ke anukUla	2
	ke bahAne	2
	kI KAwira	2
	pariNAmasvarUpa	2
	svarUpa	1
	hewu	1
	vAloM ko	1
	ke nAwe	1
	vAloM para	1
	viroXa meM	1

3.3 N-gram of tags

Include statistics about the frequency of n-gram of tags. Take n in the range [2,4]. 1. N-grams of **AnnCorra tags** are given below:

Table 31: Bi-grams of AnnCorra Tags

bi-grams	frequency	bi-grams	frequency
r6, lwg__psp	13184	lwg__psp, nmod	1868
lwg__psp, nmod__adj	9513	nmod__adj, k7t	1754
lwg__psp, lwg__psp	7827	k4, lwg__psp	1733
k7, lwg__psp	7713	lwg__psp, lwg__rp	1705
lwg__psp, pof	6378	nmod, pof__cn	1661
k1, lwg__psp	6117	k2, pof	1583
main, lwg__vaux	5621	lwg__psp, k7t	1428
ccof, lwg__vaux	5270	k1, ccof	1373
lwg__vaux, rsym	5159	k2, ccof	1282
lwg__vaux, lwg__vaux_cont	5008	lwg__neg, ccof	1218
lwg__psp, k1	4555	nmod, lwg__psp	1157
pof__cn, pof__cn	4427	rsym, ccof	1132
ccof, rsym	4361	k5, lwg__psp	1102
lwg__psp, pof__cn	4126	k1, r6	1088
lwg__vaux_cont, rsym	4122	k2, nmod__adj	1070
lwg__psp, r6	3936	lwg__psp, k7p	1062
lwg__psp, k2	3931	rsym, pof__cn	1058
k7t, lwg__psp	3894	k1s, main	1047
k7p, lwg__psp	3781	pof__cn, k2	1036
pof, main	3597	k2, k1	998
k2, lwg__psp	3322	nmod__adj, k7p	956
pof, ccof	3311	nmod__adj, ccof	953
main, rsym	3152	lwg__psp, mod	952
nmod__adj, k1	3109	k2, main	943
pof__cn, k1	2881	k1s, ccof	941
lwg__psp, main	2805	lwg__psp, vmod	910
main, k2	2764	pof__cn, rsym	909
lwg__psp, k7	2587	ccof, r6	871
nmod__adj, r6	2548	ccof, k1	857
lwg__psp, ccof	2538	pof__cn, k7	817
pof__cn, r6	2493	nmod__adj, nmod	804
nmod__adj, k2	2406	vmod, lwg__psp	803
ccof, lwg__psp	2294	k1, pof	802
pof__cn, ccof	2260	k2, r6	800
nmod__adj, nmod__adj	2258	k1, pof__cn	786
nmod__adj, pof__cn	2198	k1, k1s	782
nmod__adj, k7	2157	pof, vmod	780
r6-k2, lwg__psp	2057	lwg__rp, nmod__adj	764
k1, main	1967	rh, lwg__psp	759
rt, lwg__psp	1882	lwg__psp, nmod__k1inv	737
k1, nmod__adj	1873	mod, nmod__adj	732

Table 32: Bi-grams of AnnCorra Tags continued

bi-grams	frequency
vmod, lwg__vaux	722
lwg__vaux_cont, lwg__vaux_cont	709
k2, pof__cn	694
nmod__adj, pof	690
k7a, lwg__psp	678
lwg__psp, k1s	677
lwg__psp, r6-k2	666
pof, lwg__neg	645
lwg__psp, k4	640
ccof, k2	630
ras-k1, lwg__psp	620
pof__cn, nmod__adj	618
pof__cn, k7p	608
k1, lwg__rp	599
nmod__relc, lwg__vaux	598
ccof, ccof	589
ccof, main	579
lwg__psp, adv	579
lwg__vaux, k2	576
k2, vmod	565
rsym, nmod__adj	560
jjmod, lwg__psp	555
lwg__psp, rt	552
ccof, nmod	538
nmod, ccof	537
lwg__rp, pof	514
pof__cn, k7t	507
pof__cn, nmod	506
adv, lwg__psp	506
main, rs	497
r6, nmod__adj	487
pof, r6	484
k3, lwg__psp	461
nmod__adj, r6-k2	458
nmod, nmod__adj	453
k1, k2	450
lwg__vaux, main	450
pof, rt	442
nmod__adj, k1s	442
main, k1	420
k7t, nmod__adj	419
main, nmod__adj	417
k1, k7t	416
lwg__vaux, lwg__psp	415
lwg__neg, main	413

bi-grams	frequency
r6-k1, lwg__psp	412
nmod__adj, adv	411
nmod__k1inv, lwg__vaux	409
r6, ccof	403
r6, k2	402
k1, lwg__neg	401
ccof, k7	397
rsym, r6	390
pof__cn, k4	384
nmod__adj, rsym	377
nmod__adj, mod	375
rsym, k1	350
lwg__rp, main	348
pof, nmod__relc	342
nmod__adj, k4	337
lwg__rp, ccof	329
lwg__rp, k1	329
rsym, main	326
k7t, k1	319
k1, k7	317
vmod, nmod__adj	315
k2, k7	298
lwg__vaux, ccof	295
k2, lwg__rp	294
rs, rsym	293
k1, adv	292
r6, k1	291
pof, r6-k2	289
k1, k7p	280
jjmod, nmod	279
r6, pof__cn	277
lwg__psp, k5	272
rs, k1	271
ccof, nmod__adj	271
lwg__rp, r6	269
k2, lwg__neg	269
mod, k1	259
k2, k7t	258
lwg__vaux, k1	257
rd, lwg__psp	254
lwg__psp, rsym	253
nmod__relc, rsym	249
k4a, lwg__psp	248
lwg__psp, nmod__k2inv	246
nmod__adj, jjmod	242

Table 33: Bi-grams of AnnCorra Tags continued

bi-grams	frequency
nmod, rsym	242
k7t, r6	241
nmod__adj, rt	240
adv, pof	237
rs, nmod__adj	236
rsym, rs	235
r6, k7	234
k2, rt	234
lwg__vaux, nmod__adj	233
pof, k7	233
lwg__psp, jjmod	233
rsym, fragof	232
ccof, rs	232
nmod__adj, k5	228
main, r6	227
k1, vmod	226
k7t, pof__cn	223
rsym, nmod	222
nmod, r6	220
ccof, k7p	219
lwg__rp, k2	215
lwg__psp, rh	214
lwg__vaux_cont, k2	214
pof, k7t	213
jjmod__intf, nmod__adj	212
vmod, ccof	209
mod, rsym	209
vmod, k1	209
ccof, lwg__rp	208
lwg__vaux, rs	208
ccof, pof__cn	208
k2, k2	208
nmod__k2inv, lwg__vaux	207
nmod, k1	205
r6, r6	204
nmod__adj, lwg__rp	204
rsym, k2	201
k1s, lwg__neg	200
pof__cn, r6-k2	197
rsym, lwg__psp	196
pof, nmod	192
nmod__adj, rh	191
rsym, k7	189
k7t, lwg__rp	187
rsym, pof__redup	185

bi-grams	frequency
lwg__vaux, pof__cn	182
lwg__rp, pof__cn	181
pof__cn, k5	181
k2, k7p	181
r6, rsym	180
rsym, k7t	178
vmod, r6	176
lwg__psp, lwg__neg	176
k7p, ccof	176
k7, ccof	174
lwg__rp, k1s	173
k7, nmod__adj	171
lwg__vaux_cont, ccof	171
k1, mod	170
main, pof__cn	169
nmod__k1inv, nmod__adj	168
pof, nmod__k1inv	168
pof, k2	167
adv, ccof	167
lwg__psp, k3	167
rs, pof__cn	166
pof__cn, k7a	164
k1, rsym	161
adv, lwg__rp	161
ccof, vmod	157
lwg__psp, jjmod__intf	156
lwg__rp, lwg__psp	156
pof__cn, jjmod	154
k2, r6-k2	154
ras-k2, lwg__psp	154
r6-k2, pof	153
main, k7t	153
lwg__psp, k2s	152
lwg__vaux, r6	152
lwg__psp, k2p	152
lwg__vaux_cont, main	152
lwg__psp, ras-k1	151
lwg__psp, nmod__relc	147
k2, rsym	146
mod, k2	146
k2, adv	144
lwg__psp, r6-k1	141
k2, nmod	140
vmod, k2	139
vmod, pof__cn	138

Table 34: **Tri-grams** of AnnCorra Tags

tri-grams	frequency
lwg__vaux, lwg__vaux_cont, rsym	3527
lwg__psp, r6, lwg__psp	3090
main, lwg__vaux, rsym	2634
r6, lwg__psp, nmod__adj	2582
nmod__adj, r6, lwg__psp	2527
pof, main, lwg__vaux	2476
lwg__psp, k7, lwg__psp	2459
main, lwg__vaux, lwg__vaux_cont	2395
pof__cn, r6, lwg__psp	2375
ccof, lwg__vaux, lwg__vaux_cont	2223
pof, ccof, lwg__vaux	2220
ccof, lwg__vaux, rsym	2182
nmod__adj, k7, lwg__psp	2131
lwg__psp, pof, main	1993
pof__cn, k1, lwg__psp	1860
r6, lwg__psp, k1	1807
r6-k2, lwg__psp, pof	1803
lwg__psp, pof, ccof	1678
r6, lwg__psp, k2	1614
rt, lwg__psp, lwg__psp	1587
r6, lwg__psp, k7	1543
lwg__psp, main, k2	1366
lwg__psp, lwg__psp, nmod__adj	1324
k1, lwg__psp, main	1291
lwg__psp, nmod__adj, k1	1183
lwg__psp, nmod__adj, k2	1159
pof__cn, pof__cn, pof__cn	1156
k7, lwg__psp, nmod__adj	1131
k7t, lwg__psp, lwg__psp	1112
lwg__psp, k1, lwg__psp	1023
k1, main, k2	1010
lwg__psp, main, lwg__vaux	1005
lwg__psp, k7t, lwg__psp	1004
r6, lwg__psp, pof__cn	1002
lwg__psp, nmod__adj, r6	1001
k1, lwg__psp, nmod__adj	988
lwg__psp, nmod__adj, nmod__adj	966
nmod__adj, k2, lwg__psp	964
k7, lwg__psp, lwg__psp	952
k7, lwg__psp, pof	931
nmod__adj, k7p, lwg__psp	926
k2, lwg__psp, pof	905
lwg__psp, k2, lwg__psp	898
pof, main, rsym	895
nmod__adj, k7t, lwg__psp	882

tri-grams	frequency
lwg__psp, k7p, lwg__psp	860
nmod__adj, k1, lwg__psp	848
lwg__psp, pof__cn, pof__cn	820
ccof, rsym, ccof	811
lwg__psp, ccof, lwg__vaux	803
pof__cn, k7, lwg__psp	793
lwg__psp, lwg__psp, lwg__psp	789
lwg__psp, nmod__adj, k7	779
r6, lwg__psp, nmod	756
k2, main, lwg__vaux	720
lwg__psp, lwg__psp, pof__cn	715
k7p, lwg__psp, nmod__adj	707
lwg__vaux, lwg__vaux_cont, lwg__vaux_cont	707
lwg__psp, k2, pof	699
pof__cn, ccof, rsym	693
pof__cn, pof__cn, k1	683
lwg__psp, lwg__psp, k1	682
pof__cn, pof__cn, r6	680
lwg__psp, k1, main	679
k1, r6, lwg__psp	675
k2, ccof, lwg__vaux	671
k2, r6, lwg__psp	656
k1s, main, rsym	645
pof__cn, pof__cn, ccof	645
lwg__psp, nmod, pof__cn	642
r6, lwg__psp, r6	634
main, k2, k1	634
pof, ccof, rsym	623
k7a, lwg__psp, lwg__psp	616
lwg__psp, nmod__adj, pof__cn	609
pof__cn, ccof, lwg__psp	608
k7, lwg__psp, r6	602
lwg__vaux_cont, lwg__vaux_cont, rsym	595
lwg__psp, lwg__psp, r6	592
lwg__psp, r6-k2, lwg__psp	590
r6, lwg__psp, ccof	589
pof__cn, k7p, lwg__psp	584
k7, lwg__psp, k1	574
main, k2, nmod__adj	574
lwg__psp, nmod__adj, k7t	564
lwg__psp, pof__cn, r6	561
k1, lwg__psp, r6	557
lwg__neg, ccof, rsym	541
lwg__psp, k4, lwg__psp	536
k7t, lwg__psp, nmod__adj	525

Table 35: **Tri-grams** of AnnCorra Tags continued

tri-grams	frequency
nmod, pof__cn, pof__cn	525
lwg__psp, rt, lwg__psp	518
pof__cn, k2, lwg__psp	518
ccof, rsym, pof__cn	515
k1s, ccof, rsym	512
lwg__neg, ccof, lwg__vaux	507
pof__cn, k7t, lwg__psp	489
lwg__psp, pof, vmod	488
k7, lwg__psp, pof__cn	482
lwg__psp, k1, ccof	477
lwg__psp, lwg__psp, pof	466
k2, pof, ccof	463
lwg__psp, ccof, rsym	462
nmod__adj, pof__cn, k1	460
k2, pof, main	457
nmod__adj, r6-k2, lwg__psp	455
k1, lwg__psp, pof__cn	454
vmod, lwg__psp, lwg__psp	452
lwg__psp, k2, main	446
lwg__psp, k2, ccof	440
lwg__psp, lwg__psp, k2	439
rsym, ccof, rsym	435
pof, r6, lwg__psp	433
k7, lwg__psp, main	432
lwg__psp, nmod__adj, pof	425
lwg__psp, pof__cn, k1	424
jjmod, lwg__psp, nmod	421
lwg__psp, k1, nmod__adj	419
lwg__psp, nmod__adj, k7p	409
pof, lwg__neg, ccof	409
ccof, lwg__psp, lwg__psp	407
pof, rt, lwg__psp	403
k1, lwg__psp, k7t	401
k7, lwg__psp, k2	400
nmod__adj, nmod, pof__cn	399
pof__cn, nmod__adj, pof__cn	393
lwg__psp, lwg__rp, nmod__adj	392
lwg__psp, main, rsym	391
k7, lwg__psp, ccof	388
k1, ccof, rsym	387
nmod, pof__cn, k1	385
k1, main, rsym	383
k1, ccof, lwg__vaux	381
k1, lwg__psp, k2	381
nmod__adj, nmod__adj, k1	381

tri-grams	frequency
k7p, lwg__psp, pof__cn	380
lwg__psp, nmod, lwg__psp	380
lwg__psp, lwg__psp, ccof	379
k1, k1s, main	378
main, k2, pof__cn	375
k1, main, lwg__vaux	372
ccof, lwg__psp, nmod__adj	367
k7p, lwg__psp, r6	365
pof__cn, k4, lwg__psp	365
main, lwg__vaux, k2	362
nmod__adj, pof__cn, pof__cn	361
pof, vmod, lwg__vaux	353
lwg__psp, nmod__adj, nmod	350
lwg__psp, pof__cn, k2	350
lwg__psp, mod, nmod__adj	348
nmod__adj, nmod__adj, pof__cn	346
k2, lwg__psp, nmod__adj	343
ccof, nmod, ccof	342
k4, lwg__psp, nmod__adj	342
k7t, lwg__psp, pof__cn	337
r6-k1, lwg__psp, pof	333
k2, lwg__psp, vmod	333
nmod__adj, k4, lwg__psp	332
r6, ccof, lwg__psp	325
nmod__adj, k2, pof	324
rsym, pof__cn, pof__cn	321
ras-k1, lwg__psp, lwg__psp	320
rh, lwg__psp, lwg__psp	319
lwg__psp, lwg__psp, lwg__rp	316
nmod, lwg__psp, nmod__adj	316
lwg__psp, pof, lwg__neg	308
rsym, pof__cn, ccof	305
main, k2, r6	304
k1, lwg__psp, pof	300
nmod__adj, k1, ccof	298
nmod__adj, ccof, lwg__psp	293
lwg__psp, pof__cn, rsym	293
k7t, lwg__psp, r6	292
nmod__adj, k1, nmod__adj	289
lwg__psp, lwg__rp, pof	289
ccof, lwg__vaux, main	288
lwg__psp, nmod__adj, ccof	287
pof__cn, nmod, pof__cn	287
lwg__psp, k1, pof	286
ccof, r6, ccof	284

Table 36: **Tri-grams** of AnnCorra Tags continued

tri-grams	frequency
nmod__adj, adv, lwg__psp	284
k7p, lwg__psp, pof	282
k1, lwg__psp, k1	281
nmod, pof__cn, ccof	277
rsym, r6, lwg__psp	275
k1, pof, ccof	275
lwg__psp, k1s, ccof	272
lwg__rp, pof, main	270
r6, lwg__psp, k7p	270
pof, r6-k2, lwg__psp	270
k1, k7, lwg__psp	266
k2, k7, lwg__psp	265
lwg__psp, k5, lwg__psp	265
lwg__psp, pof, r6	265
lwg__psp, pof, rt	263
nmod__relc, lwg__vaux, rsym	260
nmod__adj, nmod__adj, k2	258
nmod__adj, k2, main	256
nmod__adj, pof__cn, ccof	256
r6, lwg__psp, k1s	255
nmod__adj, k1, main	254
k4, lwg__psp, pof	254
k1, lwg__neg, ccof	254
k1, k1s, ccof	254
k7, lwg__psp, lwg__rp	250
k7t, lwg__psp, k1	249
nmod__relc, lwg__vaux, lwg__vaux_cont	248
pof, nmod__relc, lwg__vaux	247
k1, pof, main	245
r6, lwg__psp, k7t	244
nmod__adj, rt, lwg__psp	240
nmod__adj, pof, ccof	240
rd, lwg__psp, lwg__psp	239
lwg__psp, k1s, main	238
nmod__adj, pof__cn, r6	237
ccof, lwg__vaux, ccof	237
ccof, lwg__psp, pof	234
lwg__psp, nmod, nmod__adj	233
lwg__psp, pof__cn, k7	231
nmod__adj, k2, ccof	227
nmod__adj, k5, lwg__psp	227
lwg__psp, nmod__adj, r6-k2	227
rsym, rs, rsym	224
k1, k7p, lwg__psp	224
k1s, main, rs	223

tri-grams	frequency
r6, k7, lwg__psp	221
lwg__psp, pof__cn, ccof	221
k4, lwg__psp, k2	221
k7p, lwg__psp, k2	221
k1, k7t, lwg__psp	221
pof__cn, rsym, pof__cn	218
k1, nmod__adj, k7	217
lwg__psp, k2, vmod	216
k1, nmod__adj, r6	215
k2, k1, lwg__psp	213
lwg__psp, k1, k1s	212
k7p, lwg__psp, lwg__psp	211
pof__cn, ccof, k1	210
pof, k7, lwg__psp	209
ccof, k1, ccof	208
k1, lwg__psp, k4	208
k7p, lwg__psp, k1	206
k2, rt, lwg__psp	205
r6, lwg__psp, r6-k2	203
nmod__adj, pof, main	202
lwg__psp, lwg__psp, k7t	202
lwg__neg, main, rsym	201
k7p, lwg__psp, ccof	200
lwg__psp, nmod__k1inv, lwg__vaux	198
lwg__psp, lwg__psp, main	197
lwg__rp, main, lwg__vaux	197
nmod, r6, lwg__psp	196
r6, lwg__psp, rt	194
lwg__rp, r6, lwg__psp	191
pof__cn, r6-k2, lwg__psp	191
lwg__psp, rh, lwg__psp	191
lwg__psp, vmod, lwg__psp	190
lwg__vaux, lwg__vaux_cont, k2	189
lwg__psp, r6, k2	188
lwg__psp, lwg__psp, nmod	188
ccof, r6, lwg__psp	187
nmod__adj, nmod__adj, r6	187
k1, nmod__adj, nmod__adj	186
nmod__adj, k1s, ccof	186
k2, lwg__psp, ccof	186
pof, k7t, lwg__psp	185
k2, main, rsym	185
k7t, r6, lwg__psp	185
pof__cn, pof__cn, k2	185
k7, lwg__psp, mod	185

Table 37: **Four-grams** of AnnCorra Tags

four-grams	frequency
main, lwg__vaux, lwg__vaux_cont, rsym	1828
r6, lwg__psp, k7, lwg__psp	1519
ccof, lwg__vaux, lwg__vaux_cont, rsym	1463
lwg__psp, pof, main, lwg__vaux	1393
pof, main, lwg__vaux, rsym	1296
lwg__psp, pof, ccof, lwg__vaux	1138
k1, lwg__psp, main, k2	1093
pof, main, lwg__vaux, lwg__vaux_cont	1016
lwg__psp, nmod__adj, r6, lwg__psp	995
pof, ccof, lwg__vaux, rsym	947
pof, ccof, lwg__vaux, lwg__vaux_cont	937
lwg__psp, nmod__adj, k7, lwg__psp	772
r6-k2, lwg__psp, pof, main	641
lwg__psp, r6, lwg__psp, nmod__adj	629
pof__cn, pof__cn, r6, lwg__psp	595
lwg__vaux, lwg__vaux_cont, lwg__vaux_cont, rsym	593
r6, lwg__psp, r6, lwg__psp	584
lwg__psp, pof__cn, r6, lwg__psp	557
lwg__psp, r6-k2, lwg__psp, pof	506
r6, lwg__psp, nmod, pof__cn	501
r6-k2, lwg__psp, pof, ccof	484
lwg__psp, pof, main, rsym	483
k7, lwg__psp, r6, lwg__psp	478
pof__cn, pof__cn, k1, lwg__psp	474
lwg__psp, r6, lwg__psp, k1	470
lwg__psp, rt, lwg__psp, lwg__psp	464
lwg__psp, r6, lwg__psp, k2	463
pof__cn, r6, lwg__psp, nmod__adj	460
r6, lwg__psp, k2, lwg__psp	449
lwg__psp, lwg__psp, r6, lwg__psp	446
pof__cn, k1, lwg__psp, main	440
nmod__adj, r6, lwg__psp, nmod__adj	437
r6, lwg__psp, nmod__adj, k1	431
lwg__psp, main, lwg__vaux, lwg__vaux_cont	426
lwg__psp, nmod__adj, k2, lwg__psp	409
nmod__adj, r6-k2, lwg__psp, pof	401
lwg__psp, main, lwg__vaux, rsym	400
lwg__psp, nmod__adj, k7p, lwg__psp	390
k7, lwg__psp, lwg__psp, lwg__psp	389
nmod__adj, r6, lwg__psp, k2	388
nmod__adj, r6, lwg__psp, k1	383
ccof, rsym, ccof, rsym	383
pof__cn, ccof, rsym, pof__cn	381
k2, main, lwg__vaux, rsym	374
k2, pof, main, lwg__vaux	358

Table 38: **Four-grams** of AnnCorra Tags continued

four-grams	frequency
lwg__psp, k7, lwg__psp, nmod__adj	352
lwg__psp, ccof, lwg__vaux, lwg__vaux_cont	348
pof__cn, pof__cn, pof__cn, pof__cn	342
k1, lwg__psp, r6, lwg__psp	340
pof, rt, lwg__psp, lwg__psp	337
lwg__psp, k7, lwg__psp, pof	337
ccof, lwg__vaux, lwg__vaux_cont, lwg__vaux_cont	337
lwg__psp, r6, lwg__psp, k7	336
lwg__psp, main, k2, k1	336
k2, pof, ccof, lwg__vaux	333
lwg__psp, k2, main, lwg__vaux	329
nmod__adj, k7, lwg__psp, nmod__adj	329
main, lwg__vaux, lwg__vaux_cont, lwg__vaux_cont	328
pof__cn, r6, lwg__psp, k7	327
rt, lwg__psp, lwg__psp, nmod__adj	315
lwg__psp, pof, ccof, rsym	315
k1, lwg__psp, k7t, lwg__psp	312
k7p, lwg__psp, r6, lwg__psp	309
nmod, pof__cn, k1, lwg__psp	306
lwg__psp, k7t, lwg__psp, lwg__psp	305
nmod__adj, r6, lwg__psp, k7	305
lwg__psp, ccof, lwg__vaux, rsym	301
rsym, ccof, rsym, ccof	296
k2, ccof, lwg__vaux, rsym	295
k2, main, lwg__vaux, lwg__vaux_cont	294
lwg__psp, k1, main, k2	291
lwg__psp, nmod__adj, k1, lwg__psp	287
lwg__psp, main, k2, nmod__adj	286
lwg__psp, k2, ccof, lwg__vaux	283
nmod__adj, k2, lwg__psp, pof	282
nmod__adj, pof__cn, k1, lwg__psp	281
lwg__psp, nmod__adj, k7t, lwg__psp	278
pof__cn, r6, lwg__psp, nmod	275
ccof, r6, ccof, lwg__psp	270
r6, lwg__psp, nmod__adj, nmod__adj	268
ccof, rsym, pof__cn, ccof	267
k2, ccof, lwg__vaux, lwg__vaux_cont	265
lwg__psp, pof, vmod, lwg__vaux	263
k7, lwg__psp, pof, ccof	263
lwg__psp, k7, lwg__psp, lwg__psp	261
pof__cn, k1, lwg__psp, nmod__adj	261
lwg__psp, k2, lwg__psp, pof	259
r6, lwg__psp, nmod__adj, k2	257
r6, lwg__psp, k1, main	255
nmod__adj, k7, lwg__psp, lwg__psp	254

Table 39: **Four-grams** of AnnCorra Tags continued

four-grams	frequency
pof, r6-k2, lwg__psp, pof	253
lwg__neg, ccof, lwg__vaux, rsym	252
r6, lwg__psp, k2, pof	250
k7, lwg__psp, pof, main	247
k7, lwg__psp, main, lwg__vaux	244
lwg__psp, lwg__psp, k1, lwg__psp	244
r6, lwg__psp, k1, lwg__psp	242
main, k2, r6, lwg__psp	241
nmod__adj, rt, lwg__psp, lwg__psp	239
r6, lwg__psp, k1, ccof	239
r6, lwg__psp, k7p, lwg__psp	238
nmod__adj, k7, lwg__psp, pof	237
pof__cn, pof__cn, ccof, rsym	237
lwg__psp, pof, rt, lwg__psp	236
nmod__adj, pof__cn, r6, lwg__psp	235
lwg__psp, pof, r6, lwg__psp	233
lwg__psp, pof__cn, k1, lwg__psp	230
k7t, lwg__psp, lwg__psp, nmod__adj	228
lwg__psp, pof__cn, k7, lwg__psp	227
k1, main, k2, nmod__adj	226
pof__cn, r6, lwg__psp, k1	226
lwg__psp, nmod__adj, r6-k2, lwg__psp	226
nmod__adj, k7t, lwg__psp, lwg__psp	226
k1, lwg__psp, k1, main	221
k1, main, k2, k1	219
k7, lwg__psp, k1, lwg__psp	218
pof, lwg__neg, ccof, lwg__vaux	216
k1, nmod__adj, r6, lwg__psp	214
k1, nmod__adj, k7, lwg__psp	214
r6, lwg__psp, nmod__adj, k7	213
lwg__psp, pof__cn, pof__cn, pof__cn	211
k7t, lwg__psp, r6, lwg__psp	210
pof__cn, r6, lwg__psp, k2	210
pof__cn, r6, lwg__psp, pof__cn	209
lwg__psp, r6, lwg__psp, pof__cn	209
k2, lwg__psp, pof, ccof	209
lwg__psp, k2, pof, ccof	209
lwg__psp, k2, pof, main	208
rsym, pof__cn, ccof, rsym	205
lwg__psp, k1, lwg__psp, nmod__adj	202
r6, lwg__psp, r6-k2, lwg__psp	201
k1, k1s, main, rs	201
lwg__psp, nmod__adj, nmod, pof__cn	201
lwg__psp, nmod, pof__cn, pof__cn	200
pof__cn, pof__cn, pof__cn, r6	199

Table 40: Bi-grams of UD Tags

Bi-grams	frequency
obl, case	18561
nmod, case	15366
case, compound	8829
case, case	6818
case, nmod	6677
case, obl	6161
nsubj, case	5786
compound, nmod	4956
case, amod	4822
root, aux	4796
aux:pass, punct	4635
compound, compound	4310
compound, root	4169
aux, punct	4148
case, obj	3965
case, nsubj	3503
compound, obl	3375
punct, nmod	3295
punct, nsubj	3191
punct, obl	3168
obj, case	3067
case, root	3059
compound, nsubj	3042
punct, compound	3002
root, mark	2784
aux, aux:pass	2746
amod, obl	2690
nmod, compound	2681
aux:pass, aux:pass	2616
compound, obj	2431
amod, nmod	2329
root, punct	2115
case, nummod	2063
case, det	2021
conj, case	1943
mark, compound	1927
cop, punct	1898
compound, conj	1886
mark, nsubj	1873
nsubj, root	1803
det, obl	1795
mark, mark	1699
obj, compound	1620
root, aux:pass	1619
iobj, case	1613

Bi-grams	frequency
nmod, mark	1550
nmod, nmod	1490
case, dep	1466
mark, nmod	1449
obj, aux	1446
cc, conj	1438
nsubj, nmod	1435
compound, advcl	1409
amod, compound	1401
nsubj, obl	1329
nmod, obl	1309
mark, obl	1287
nsubj, compound	1278
amod, obj	1270
root, cop	1261
punct, det	1193
advcl, mark	1138
nummod, obl	1137
obj, root	1137
case, acl	1136
nmod, punct	1134
amod, nsubj	1113
nummod, nmod	1078
conj, punct	1010
punct, amod	991
det, nmod	984
advcl, aux	927
compound, punct	866
case, advcl	855
nmod, cc	855
cc, compound	842
cc, nmod	836
aux, mark	836
obj, punct	834
obl, mark	829
det, nsubj	803
det, obj	790
nsubj, obj	763
aux:pass, mark	761
punct, cc	759
nmod, nsubj	750
case, advmod	736
mark, obj	722
nsubj, amod	706
mark, amod	684

Table 41: Bi-grams of UD Tags continued

Bi-grams	frequency
conj, aux	680
obj, advcl	668
dep, compound	659
conj, cc	649
obl, compound	635
acl, aux	632
compound, advmod	626
obl, nmod	625
mark, det	625
nummod, amod	618
punct, conj	613
obl, cc	605
nsubj, det	589
obj, nmod	584
nmod, obj	573
case, iobj	558
obj, aux:pass	555
advmod, root	551
obl, amod	541
nsubj, dep	526
obl, obl	523
obj, obj	510
compound, acl	497
cop, mark	497
cc, amod	496
advmod, obj	492
obl, nsubj	491
obj, advmod	478
nsubj, advmod	475
case, conj	475
punct, mark	475
obj, obl	471
amod, mark	465
amod, amod	457
nummod, nsubj	457
dep, root	445
acl, aux:pass	440
acl:relcl, aux	430
det, amod	424
nummod, obj	424
dep, nmod	415
compound, iobj	408
cc, obl	404
obj, cop	391
punct, obj	390

Bi-grams	frequency
nsubj, cc	389
cc, nsubj	382
obl, root	378
nmod, amod	376
amod, conj	372
punct, nummod	365
aux, cc	365
obj, cc	350
compound, acl:relcl	350
advmod, cop	350
det, compound	345
obl, punct	345
dep, obl	342
aux:pass, cc	326
dep, nummod	326
nsubj, nummod	305
dep, obj	292
nsubj, nsubj	292
case, punct	288
nsubj, punct	284
obl, obj	283
dep, nsubj	278
obl, dep	273
obj, dep	266
mark, root	265
nsubj:pass, case	264
punct, root	254
case, cop	253
conj, aux:pass	253
nummod, compound	252
punct, punct	248
advmod, compound	247
xcomp, root	245
amod, nummod	242
root, cc	239
punct, dislocated	239
advmod, amod	238
mark, nummod	237
compound, amod	227
dep, amod	224
mark, dep	217
dep, det	214
aux, nsubj	213
obj, acl	208
acl, obl	206

Table 42: Bi-grams of UD Tags continued

Bi-grams	frequency
obj, mark	205
advmod, conj	199
obj, amod	198
nsubj, advcl	197
cop, cc	194
acl, nmod	194
conj, compound	194
punct, case	193
cc, det	193
acl, punct	192
amod, root	191
obl, det	188
obj, conj	188
conj, mark	187
nsubj, xcomp	182
advcl, nmod	178
amod, iobj	176
aux, obl	176
cc, nummod	175
nmod, conj	174
nmod, nummod	169
acl:relcl, aux:pass	169
advcl, compound	167
advcl, obl	166
advmod, nmod	162
aux, compound	161
obj, nsubj	158
root, case	155
aux:pass, obl	154
advmod, dep	153
aux, nmod	150
nmod, dep	150
obl, advmod	150
aux, root	150
amod, cc	149
conj, cop	147
case, acl:relcl	146
obl, nummod	146
mark, iobj	146
conj, nmod	144
root, advmod	144
mark, punct	143
acl, root	141
advmod, acl	140
acl, compound	138

Bi-grams	frequency
nsubj, aux	138
advcl, obj	137
obl, advcl	136
dep, dep	132
acl:relcl, punct	131
advmod, advcl	131
det, nummod	131
nsubj, conj	129
nmod, aux:pass	128
case, xcomp	126
det, root	126
dep, case	126
conj, obl	125
conj, dep	123
obl, acl	122
nmod, det	121
cc, obj	121
compound, dep	114
obj, acl:relcl	113
compound, nsubj:pass	111
punct, iobj	111
det, dep	110
amod, det	110
det, det	110
nsubj, iobj	108
advmod, obl	106
conj, root	105
advcl, aux:pass	105
advcl, dep	104
advcl, root	103
amod, punct	103
dep, advmod	103
nsubj, acl:relcl	101
aux:pass, nmod	100
acl, amod	100
nummod, conj	99
nmod, iobj	98
mark, advmod	96
advcl, nsubj	92
nmod, cop	91
case, nsubj:pass	91
punct, advmod	90
acl:relcl, cop	89
acl, obj	87
advmod, det	87

Table 43: **Tri-grams** of UD Tags

Tri-grams	frequency
case, obl, case	4877
obl, case, case	4600
case, nmod, case	3633
obl, case, compound	2825
compound, nmod, case	2761
nmod, case, compound	2625
nmod, case, obl	2574
compound, obl, case	2556
nmod, case, nmod	2436
root, aux, punct	2432
amod, obl, case	2259
aux, aux:pass, punct	2236
compound, root, aux	2151
punct, obl, case	2117
obl, case, nmod	1924
case, compound, root	1782
compound, nsubj, case	1751
punct, nmod, case	1675
aux:pass, aux:pass, punct	1651
root, aux, aux:pass	1603
nmod, case, amod	1587
amod, nmod, case	1579
obl, case, amod	1539
nmod, case, obj	1508
nmod, case, nsubj	1403
case, compound, nmod	1380
det, obl, case	1335
obl, case, obl	1323
punct, nsubj, case	1321
case, root, mark	1321
nsubj, case, root	1262
case, amod, obl	1177
root, aux:pass, aux:pass	1102
compound, compound, compound	1091
nmod, obl, case	1074
compound, compound, nmod	1065
obl, case, obj	1051
obl, case, nsubj	1044
case, compound, obl	1042
case, amod, nmod	1001
nsubj, root, mark	990
case, case, compound	982
case, compound, obj	955
obj, case, compound	937
compound, root, punct	937

Tri-grams	frequency
mark, obl, case	935
aux:pass, punct, obl	931
aux:pass, punct, nsubj	926
nmod, nmod, case	915
det, nmod, case	896
nummod, obl, case	871
aux, punct, nsubj	863
case, nsubj, case	858
nsubj, case, obl	858
mark, nmod, case	856
obl, case, root	853
cc, conj, case	852
nsubj, obl, case	837
case, case, case	834
advcl, mark, mark	810
nsubj, case, nmod	807
case, compound, compound	806
case, case, nmod	801
case, obj, case	791
aux, punct, obl	788
root, cop, punct	779
case, compound, advcl	775
aux:pass, punct, nmod	769
punct, compound, compound	761
nmod, case, case	753
case, root, aux	750
case, nmod, compound	742
nmod, compound, nsubj	736
case, case, amod	716
nsubj, nmod, case	715
nmod, compound, compound	715
obj, aux, punct	712
nsubj, case, compound	711
compound, compound, nsubj	708
nummod, nmod, case	690
root, mark, nsubj	690
aux, punct, nmod	683
compound, nmod, mark	683
punct, nsubj, root	683
nmod, case, nummod	680
compound, root, aux:pass	678
punct, compound, nmod	664
case, amod, compound	650
compound, nmod, compound	650
aux:pass, punct, compound	640

Table 44: **Tri-grams** of UD Tags continued

Tri-grams	frequency
case, obj, compound	628
obl, case, acl	627
compound, obj, aux	615
obl, case, nummod	614
nmod, mark, compound	607
obl, case, dep	603
case, amod, obj	590
obl, case, det	578
case, case, obl	568
aux:pass, aux:pass, aux:pass	568
compound, compound, obl	567
nmod, case, det	545
obj, root, aux	545
compound, conj, case	543
punct, det, obl	533
root, mark, nmod	529
obj, aux, aux:pass	528
case, nsubj, root	527
obj, compound, root	523
compound, advcl, mark	522
case, case, nsubj	514
aux, punct, compound	504
case, nummod, nmod	503
compound, obj, case	497
mark, compound, root	495
case, case, obj	495
case, det, obl	482
case, nmod, obl	481
root, aux, mark	474
root, punct, obl	464
punct, compound, nsubj	463
compound, conj, punct	460
case, compound, nsubj	458
case, iobj, case	455
root, punct, nsubj	445
case, case, nummod	434
amod, obj, case	433
punct, nmod, punct	430
case, amod, nsubj	428
case, nummod, obl	427
case, obj, root	424
case, nmod, nmod	422
root, aux:pass, punct	422
nmod, cc, conj	419
compound, advcl, aux	419

Tri-grams	frequency
cc, nmod, case	415
obl, mark, mark	410
nsubj, root, aux	404
conj, aux, punct	400
mark, nsubj, case	397
obj, case, advcl	394
cop, punct, nsubj	389
compound, obl, mark	388
aux:pass, punct, det	385
nsubj, case, obj	384
root, mark, compound	383
punct, compound, obl	381
obj, aux:pass, aux:pass	379
cc, compound, conj	375
obj, nmod, mark	374
root, mark, obl	373
cop, punct, nmod	372
obj, case, obl	368
nmod, compound, nmod	365
det, obj, case	363
compound, compound, conj	361
amod, nsubj, case	360
conj, case, case	357
aux, punct, det	357
compound, iobj, case	356
amod, nmod, compound	356
case, det, obj	355
root, cop, mark	351
cop, punct, obl	348
iobj, case, compound	348
nsubj, case, amod	346
nsubj, case, det	340
obl, obl, case	338
root, punct, nmod	338
case, det, nmod	328
case, root, cop	325
amod, compound, nmod	324
compound, obj, punct	323
case, nsubj, compound	317
case, dep, compound	317
nmod, case, root	314
case, obj, advcl	310
case, nummod, amod	308
case, compound, advmod	307
punct, nmod, nmod	304

Table 45: **Tri-grams** of UD Tags continued

Tri-grams	frequency
obl, case, advcl	303
obl, nmod, case	300
root, punct, compound	297
compound, nmod, punct	296
dep, compound, root	295
compound, conj, aux	295
punct, conj, cc	291
nsubj, case, nsubj	289
conj, punct, compound	287
advmod, cop, punct	287
conj, case, compound	286
compound, punct, compound	284
case, root, aux:pass	281
nmod, punct, nmod	281
case, nsubj, obl	279
nmod, nmod, compound	278
nmod, case, conj	277
case, case, det	276
punct, amod, nmod	275
case, compound, acl	275
mark, nsubj, nmod	274
case, compound, conj	272
mark, compound, nmod	272
obl, cc, conj	271
obj, root, punct	270
punct, amod, obl	269
aux:pass, punct, amod	269
nmod, nsubj, case	268
nsubj, compound, root	267
punct, conj, punct	265
nmod, mark, obj	265
case, compound, punct	264
case, det, nsubj	264
case, nummod, obj	263
obj, case, nmod	262
obj, cop, punct	261
cc, obl, case	260
aux, punct, cc	260
conj, punct, conj	257
root, mark, det	257
case, nmod, mark	255
nsubj, amod, obl	253
mark, mark, compound	253
mark, mark, nsubj	252
mark, nsubj, obl	252

Tri-grams	frequency
case, case, dep	251
case, nmod, cc	250
case, nsubj, nmod	250
case, obj, nmod	249
compound, obj, aux:pass	249
iobj, case, obj	247
obj, root, aux:pass	246
obj, compound, advcl	242
punct, nmod, compound	241
dep, nmod, case	241
punct, compound, conj	240
conj, cc, conj	239
advmod, root, aux	239
case, root, punct	239
case, case, root	235
punct, nsubj, nmod	233
nmod, compound, obl	232
case, acl, aux:pass	230
obl, case, advmod	229
case, nummod, nsubj	226
nsubj, compound, obl	225
nmod, compound, conj	224
aux, punct, amod	222
mark, mark, nmod	222
acl:relcl, aux, punct	221
conj, case, obl	221
amod, conj, case	219
compound, acl, aux	219
aux, mark, nsubj	219
compound, conj, cc	218
case, nmod, obj	218
dep, obl, case	217
aux:pass, aux:pass, mark	217
compound, punct, nmod	217
cc, compound, compound	216
mark, compound, nsubj	213
compound, compound, obj	211
compound, root, mark	211
root, aux, cc	208
obj, punct, nsubj	208
root, mark, amod	208
amod, compound, obl	207
obj, advcl, mark	207
nmod, punct, dislocated	207
nsubj, det, obl	207

Table 46: **Four-grams** of UD Tags

Four-grams	frequency
nmod, case, obl, case	2306
root, aux, aux:pass, punct	1343
nmod, case, nmod, case	1331
case, obl, case, case	1211
compound, root, aux, punct	1189
obl, case, nmod, case	1073
nsubj, case, root, mark	1054
case, amod, obl, case	1003
obl, case, obl, case	886
case, compound, root, aux	882
case, obl, case, compound	816
compound, obl, case, case	735
compound, root, aux, aux:pass	724
case, nmod, case, compound	706
case, compound, obl, case	703
obl, case, case, compound	689
case, amod, nmod, case	669
nsubj, case, obl, case	663
obl, case, case, case	662
compound, compound, nmod, case	654
root, aux:pass, aux:pass, punct	633
aux:pass, punct, obl, case	611
case, compound, nmod, case	606
amod, obl, case, case	597
punct, nsubj, root, mark	587
compound, nmod, case, nmod	575
compound, nmod, case, obl	565
case, nmod, case, obl	561
obl, case, compound, root	556
obl, case, case, nmod	537
aux, punct, obl, case	530
nmod, case, compound, root	523
punct, obl, case, case	508
nmod, compound, nsubj, case	499
nmod, case, nmod, compound	497
punct, nsubj, case, root	490
aux, aux:pass, punct, nsubj	484
case, case, nmod, case	473
case, nmod, case, nmod	473
aux:pass, aux:pass, aux:pass, punct	472
root, aux, punct, obl	467
case, obl, case, nmod	461
obl, case, compound, nmod	459
compound, compound, obl, case	457
compound, compound, nsubj, case	453

Table 47: **Four-grams** of UD Tags continued

Four-grams	frequency
root, aux, punct, nsubj	452
aux:pass, punct, nmod, case	448
case, nmod, case, obj	445
compound, root, aux:pass, aux:pass	442
obl, case, case, obl	441
compound, nsubj, case, root	432
root, aux, punct, nmod	430
obl, case, case, amod	429
nmod, case, compound, nmod	428
aux:pass, punct, nsubj, case	421
nmod, case, obj, case	419
obj, aux, aux:pass, punct	416
aux, aux:pass, punct, obl	414
compound, nmod, case, compound	412
nmod, case, amod, nmod	410
punct, compound, nmod, case	400
nsubj, case, nmod, case	399
case, nmod, obl, case	398
compound, advcl, mark, mark	397
case, case, obl, case	392
case, compound, root, punct	392
case, obl, case, amod	392
case, nmod, case, amod	387
aux, punct, nmod, case	386
aux, aux:pass, punct, nmod	380
obl, case, root, aux	378
case, det, obl, case	378
nmod, case, amod, obl	372
amod, obl, case, compound	371
obl, case, nsubj, case	371
punct, det, obl, case	370
compound, obl, case, compound	360
case, compound, nmod, mark	360
aux, punct, nsubj, case	352
punct, nmod, case, nmod	348
case, root, mark, nsubj	344
obl, case, case, nsubj	342
aux:pass, aux:pass, punct, obl	339
compound, nsubj, case, obl	337
case, nummod, obl, case	335
case, nmod, case, nsubj	331
case, nummod, nmod, case	328
punct, compound, obl, case	327
case, compound, root, aux:pass	327
nmod, mark, compound, root	325

Table 48: **Four-grams** of UD Tags continued

Four-grams	frequency
case, obl, case, obl	324
obl, case, amod, obl	322
nmod, compound, compound, nsubj	322
punct, compound, nsubj, case	321
root, aux, punct, compound	320
obl, case, compound, obj	320
compound, compound, compound, nmod	318
case, root, aux, punct	316
root, punct, obl, case	316
obj, root, aux, punct	316
compound, compound, compound, compound	315
mark, compound, root, aux	313
root, aux:pass, aux:pass, aux:pass	313
obl, case, case, obj	312
compound, obj, aux, punct	312
root, mark, nmod, case	310
amod, nmod, case, compound	310
aux:pass, aux:pass, punct, nsubj	307
case, det, nmod, case	305
obj, compound, root, aux	303
aux, aux:pass, punct, compound	301
case, compound, advcl, mark	296
case, obl, case, obj	294
obl, case, compound, obl	293
obl, case, amod, nmod	293
punct, obl, case, compound	293
nmod, case, compound, obl	291
nmod, cc, conj, case	289
case, nmod, nmod, case	281
compound, nmod, mark, compound	280
obl, case, compound, compound	280
case, nsubj, root, mark	280
nmod, compound, nmod, case	279
nmod, case, compound, obj	276
aux:pass, aux:pass, punct, nmod	275
compound, nmod, case, amod	270
compound, obl, case, nmod	270
case, compound, obj, aux	269
root, mark, obl, case	266
obl, case, case, nummod	261
nmod, obl, case, case	260
nummod, obl, case, case	257
case, obj, case, compound	256
amod, nmod, case, obl	254
punct, compound, compound, compound	254

4 Error Analysis for automatic tagging

4.1 Errors in AnnCorra Tagging

1. The use of case markers to determine the tag causes problems especially for those case markers which are associated with a wide range of tags. The model tends to depend highly on the syntactic cues and overlooks any other cues that tend be on the semantic side of the spectrum. FOr example, the use of 'se' marker for 'ras-k*' is very common in Hindi, but the model fails to identify 'ras-k*' tag repeatedly whenever it occurs with 'se'.

   (1) rAma **SAha rUKa KAna se** milA.
   ram shah rukh khan ASSOCIATIVE meet.PERF
   'Ram met Shah Rukh Khan.'

2. The model confuses between 'nmod' and 'nmod__adj' and when they occur together it switches the order of the two tags.

   (2) **saMbaMXiwa kRewrIya** xeSa
   related regional countries
   'related regional countries'

   The model marks *saMbaMXiwa* as 'nmod' and *kRewrIya* as 'nmod__adj' instead of the correct annotation which is exactly reverse.

3. The model fails to differentiate between verbs/verb modifiers and postpositions when the verb is part of the LWG of the postposition. For example:

   (3) vicAra **kie bagEra**
   thought without
   'related regional countries'

   (4) KAna **ke havAle se**
   Khan INS
   'depending on Khan'

   In the fourth example, the model marks 'KAna' as 'r6' depending on 'ke' without looking ahead to realise the entire LWG.

4. The model fails to differentiate between 'k7' and 'k7p' when there isn't an explicit difference in the meaning of the words but there is in context. For example:

   (5) rAma **amarIkA** gayA
   ram america go.PERF
   'Ram went to America.'

   (6) merA mana **amarIkA** meM hE
   my heart america in be.PRES
   'I am mentally in America.'

   The model tags 'amarIkA' as 'k7p' correctly in example 5, but fails to understand the abstractness of 'amarIkA' in example 6, marking it incorrectly as 'k7p' instead of 'k7'.

5. Surprisingly, the model tags 'Ora' incorrectly most number of times without showing any specific pattern to its tagging.

6. Instances of 'ADJ-NOUN-*ke*' are marked incorrectly as the model marks: 'k1-r6-psp' instead of 'nmod-r6-psp' considering only NOUN with the case marker.
7. 'k1s' is incorrectly marked as 'k1'.

   (7) maMxira eka XArmika **sWala** hE
   temple one religious place be.PRES
   'Temple is a religious place.'

   The model tags 'sWala' as k1 instead of 'k1s'.
8. In case of multiple entities being joined by a conjunction, the model forgets the conjunct relation of all entities except the first two.

   (8) inake nAma rAma, SyAma **Ora** gopAla hEM.
   their namesram, shyam, gopal be.PRES.PL
   'Their names are Ram, Shaym and Gopal.'

   In this example, the model marks Ram correctly as 'ccof', Shyam as 'ccof' but when it reaches Gopal it marks it as 'k1s' instead of 'ccof' while marking 'Ora' as 'root' instead of 'k1s'.

4.2 Possible solutions

1. Since the examples 1 and 4 suggest, the model already has a lot of syntactic cues in the form of features. Adding more features to the model that fall on the semantic side of the spectrum would improve the model.
2. Since the model seems to forget the longer dependencies as seen in example 3 and 8, the model would perform better by using tools and techniques that claim to provide infinite history.

5 Discussion

5.1 Comparison of UD and AnnCorra

1. UD is a coarser tagset while AnnCorra is finer. For example: AnnCorra tags 'k3', 'k4', 'k5', 'k7', 'k7t', 'k7p' are all equivalent to 'obl' tag of UD tagset.
2. UD identifies tags syntactically while AnnCorra identifies its tags using both syntax and semantics hence, it is called a syntactico-semantic tagset. Semantics plays a major role in the AnnCorra tagset as all its tags have been formed keeping meaning of words and how they interact in mind.
3. UD tagset has its roots in the basic structure of language and AnnCorra is highly inspired by Paninian Grammar.
4. AnnCorra tagset is highly hierarchical unlike UD.

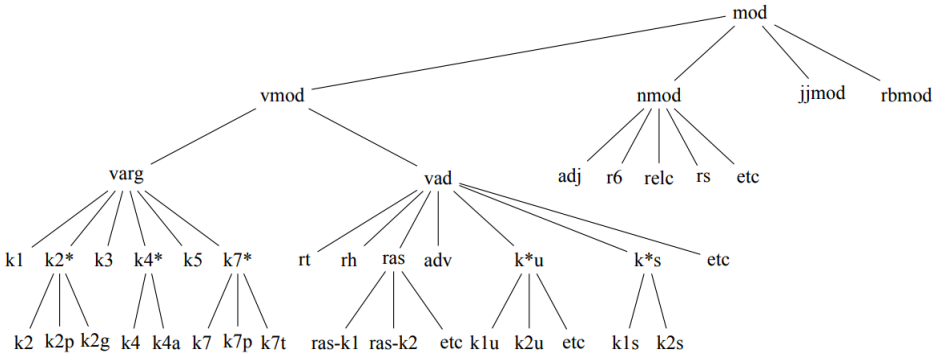

Figure 1: Hierarchy in AnnCorra

5. While working with Indian languages, AnnCorra seems to fit more perfectly as compared to UD because AnnCorra was especially designed keeping all different plays of structure and semantics in Indian Languages. Since Indian lanuages are morhologically rich and tend to have free word order, the syntactic subject-object positions cannot entirely capture the linguistics typology.
6. As seen in Section 3, AnnCorra tags have a highly exhaustive set of case markers associated with them unlike UD tags.
7. While the datasets continue to look different, one can also find some entailment between the two tagsets - hence find on-to-many mappings from UD to AnnCorra and vide versa.
8. At finer level, even if equivalent tags exist in the two tagsets, the way they mark certain types of words and linguistics phenomenon still differ. For example:

   (1) SAha rUKa KAna kA Gara suMxara hE.
   shah rukh khan GEN house beautiful be.PRES
   'Shah Rukh Khan's house is beautiful.'

   In UD, the first word in the compound name, 'Shah', becomes the head and the rest its dependents while in AnnCorra, 'Khan' is annotated the head and 'Shah' and 'Rukh' its dependents.
9. In Ellipsis, AnnCorra introduces a NULL token for the missing entity while UD doesn't.
10. In UD, the copula is not considered the head of copula construction instead the predicative nominal in the construction is annotated as the head. Whereas, in AnnCorra copula is annotated as head.
11. AnnCorra marks both coordinating and subordinating conjunctions as 'ccof' while UD separates the two as 'cc' and 'mark' respectively.
12. Adding to the previous point, in coordinating and subordinating conjunctions, UD marks the first element of conjunct as the head whereas AnnCorra marks the conjunction as head.

5.2 Need for new tags

As noticed in Section 1, both UD and AnnCorra tags take a number of case markers and hence make those markers exhaustive in nature. There is a need for finer tags to distinguish between more granular cases of grammar and language structure and to resolve some ambiguities that remain with the current set of tags.

A source of ambiguity can be seen from the example given below:

(2) rAma mohana para cIKZ uTA
ram mohan at yell.HAB rise.PERF
'Ram yelled at Mohan.'

Here, 'rAma' can be looked at from two viewpoints:

1. 'rAma' is the source of the yell, hence marked as k5.
2. 'rAma' is the agent of the action, hence marked as k1.

5.2.1 Intra-Chunk

New tag to resolve the following ambiguity:

(1) mAhI **kisi se** pUcawI hE.
mahi someone ask.HAB be.PRES
'Mahi asks someone.'

Here, 'kisi se' can be looked at from two viewpoints:

1. 'kisi' is the source of information/answer, hence marked as k5.
2. 'kisi' is the recipient of the question, hence marked as k4.

So, a new tag - **'k45'** can be introduced to justify verbs like 'mAzgA', 'pUca' etc. that can have two possible actions hidden in them.