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

Course Project Report

Instructor: Dr. Parameshwari   *Name:* Zubair Abid, *Roll Number:* 20171076

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

- Please typeset your submissions in LaTeX. Use the template provided for your answers. Please include your name and Roll number with submission.

- This report is due on 27 April at 11:55 PM.

- No extensions will be given for the submission under any circumstances.

- Submissions via any other method other than Moodle submission will be deemed invalid.

- Plagiarism of any sort will not be tolerated. Strict action will be taken for those caught in plagiarism.

1 Tags

Discussing the results of AnnCorra and UD tagging in my chosen Indian Language, Bengali.

1.1 UD

Based on the results of tagging 100 sentences with UD tags, we present the following results.

It is to be noted that a number of issues arise from tying to tag bengali with UD tags, including but not limited to a dearth of tags for specific use cases, requiring the use of a general tag covering all the various nuanced use cases that are present in the language. A detailed analysis will be provided later.

Tag	Examples	Linguistic Cues
nsubj	**bijepi** korbe kal shukrobar	POS: Noun
	ekti porjay porjonto **tader** shekhano hoy	Suffixes: -er, -r (possessive),
	tini tader byaboharic ingreji ou gonit poraten	POS: Pronoun
obj	tini **tader** byaboharic ingreji ou gonit shekhan	POS: Pronoun; Suffixes: der
	tar biddaloyer nirdhishto kono **pathshuchi** ba boi nei	POS: Noun
iobj	ekta **porjay** porjonto tader shekhano hoy	Multiple Example 2
	Multiple Example 1	Multiple Example 2
csubj	Multiple Example 1	Multiple Example 2
	Multiple Example 1	Multiple Example 2
ccomp	Multiple Example 1	Multiple Example 2
	Multiple Example 1	Multiple Example 2
ccomp	Multiple Example 1	Multiple Example 2
	Multiple Example 1	Multiple Example 2

ccomp	Multiple Example 1	Multiple Example 2
	Multiple Example 1	Multiple Example 2
ccomp	Multiple Example 1	Multiple Example 2
	Multiple Example 1	Multiple Example 2
ccomp	Multiple Example 1	Multiple Example 2
	Multiple Example 1	Multiple Example 2

Table 1: Discussion on UD tags

1.2 AnnCorra

Do the same for AnnCorra.

Tag	Examples	Linguistic Cues
k1	**bulabulite** dhAna kheYechhe	POS: Noun; Suffixes: -te
	AmAra shIta karachhe	POS: Pronoun, Suffixes: -Ara
	AmAdera siTi kaleja khuba bhAla	POS: Pronoun; Suffixes: -r (possessive)
	AnandamaTha **baNkimachandra** kartRRika rachita	POS: Noun; Postposition: kartRRika (also dbArA and diYe)
k2	bhUmikampa sAjAno gochhAno **shaharaTA** dhbansa karala	POS: Noun; Suffixes: -TA
	tini **buddhadebake** parameshbarera abatAra balena	POS: Noun; Suffixes: -ke
k2m	Ami mAke **chiThi** likhachhi	POS: Noun
k2g	Ami **mAke** chiThi likhachhi	POS: Noun; Suffixes: -ke
k3	gAdhAke **chAbuka** mAro	POS: Noun
	gAdhAke **chAbuka** diYe mAro	POS: Noun; Postposition: diYe
k4	rAm mohanke mArlo	POS: Noun; Suffixes: -ke
k5	**nadIra** ghATa theke ghaTa bhese ela	POS: Noun; Postposition: theke
	AmAra **ghara** theke mandirera chU.DA dekhA yAYa.	POS: Noun; Postposition: theke
k7p	**bADite** phulera gAchha Achhe	POS: Noun; Suffixes: -te
	kArkhAnAy kAj hoChe	POS: Noun; Suffixes: -Ay
k7t	**bhore** sUrya oThe.	POS: Noun; Suffixes: -e
	ChotobelAy luDo kHeltAm	POS: Noun; Suffixes: -Ay
k7	tArA **sAhitye** paNDita	POS: Noun; Suffixes: -ye
	tArA khuba **sukhe** Achhe	POS: Noun; Suffixes: -e
rh	**bhaYe** bhule yAYa debatAra nAma.	POS: Noun; Suffixes: -e
ru	ratana **unnatira** janya kaThora parishrama kare	POS: Noun, Suffixes: -ra, Postposition: janya
r6v	**rAmera** ekaTA meYe Achhe	POS: Noun; Suffixes: -era (genitive marker); Verb: Achhe

des	rAma bA.Di **giYechhila**	POS: Verb; Suffixes: -echhila (nominative marker)
ras	Filler text	POS: Noun; Suffixes: -e
	Filler text	POS: Noun; Suffixes: -e
k*u	Filler text	POS: Noun; Suffixes: -e
	Filler text	POS: Noun; Suffixes: -e
k*s	Filler text	POS: Noun; Suffixes: -e
	Filler text	POS: Noun; Suffixes: -e
relc	Filler text	POS: Noun; Suffixes: -e
	Filler text	POS: Noun; Suffixes: -e
adv	Filler text	POS: Noun; Suffixes: -e
	Filler text	POS: Noun; Suffixes: -e
adj	Filler text	POS: Noun; Suffixes: -e
	Filler text	POS: Noun; Suffixes: -e

Table 2: Discussion on AnnCorra tags

2 Linguistic Challenges with Annotation

(a) Differential Object marking
(b) Non-Nominative Subjects
(c) Complex Predicates
(d) Non-finite clauses: Conditional, Concessive, Relative, participial clauses
(e) Ambiguity (Coordination, Attachment)
(f) Ellipsis
(g) Non-projectivity Ex 1: I saw a man yesterday who was singing Ex 2: A hearing is scheduled on the issue today Ex 3: To his wife, John gave a fantastic gift Ex 4: Which house, John bought?
(h) Particles

3 Tag Statistics

3.1 Tag and Markers

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

There are a total of 100 k1 tags. k1 comes with the markers $0, -ne$. 0 marker is responsible for 56 of the cases and $-ne$ is responsible for the remaining 44.

3.2 Markers and Tag

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

There are a total of 100 tokens with $-ne$ marker. These tokens are marked with $k1, k2$. $k1$ marker is responsible for 32 of the cases and $k2$ is responsible for the remaining 68.

3.3 N-gram of tags

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

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

4 Error Analysis for automatic tagging

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

5 Discussion

5.1 Comparison of UD and AnnCorra

5.2 Need for new tags

5.2.1 Intra-Chunk

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

5.2.2 Inter-Chunk

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