Computational Linguistics

Assignment 1

Abhigyan Ghosh   20171089

Note: Each row in a table indicated the possible tags for a given word and each highlighted tag indicates that it is the appropriate tag for the word in that context.

Task 1:

English Development Data

She	PPS				
has	HVZ	NIL	HVZ-NC		
been	BEN				
absent	JJ	VB			
since	IN	CS	RB		
last	AP	VB	NN	AP-HL	RB
Wednesday	NR				

It	PPS	PPO-TL	PPS-TL				
doesn't	DOZ*						
matter	NN	VB					
what	WDT	WDT-HL	WDT-NC				
excuse	NN	VB					
he	PPS	PPS-NC	NIL				

gives	VBZ					
me	PPO	FW-PPO				
I	PPSS	NN	PPSS-NC	NN-TL	NP	NIL	PPSS-HL
cant	MD*					
forgive	VB					
him	PPO	PPO-NC				

I	PPSS	NN	PPSS-NC	NN-TL	NP	NIL	PPSS-HL
canceled	VBN	VBD				
my	PP$	UH				
appointment	NN					
because	CS	RB	CS-HL			
of	IN	IN-TL	IN-HL	IN-NC	IN-TL-HL	NIL
urgent	JJ					
business	NN	NN-HL				

What	WDT	WDT-HL	WPS-TL	WPO-TL	WPS		
do	DO	DO-HL	DO-NC				
you	PPSS	PPO	PPO-HL	PPSS-HL	PPSS-NC	PPO-NC	
do	DO	DO-HL	DO-NC				
in	IN	RP	IN-HL	IN-TL	IN-NC	RP-HL	NN

	NIL	RP-NC	FW-IN				
Japan	NP	NP-TL					

The	AT	AT-TL	AT-HL				
Handmaid's	NOT_FOUND						
Tale	NN						
is	BEZ	BEZ-HL	NIL	BEZ-NC	BEZ-TL		
an	AT	AT-HL	NIL	AT-NC	CC		
awesome	JJ						
piece	NN	FW-NN	VB				
of	IN	IN-TL	IN-HL	IN-NC	IN-TL-HL	NIL	
dystopian	JJ						
fiction	NN						

OK	JJ	RB	
Now	RB	RB-HL	RB-NC
what	WDT	WDT-HL	WDT-NC

I	PPSS	NN	PPSS-NC	NN-TL	NP	NIL	PPSS-HL
was	BEDZ	BEDZ-HL	BEDZ-NC				
laughed	VBD	VBN					
at	IN	IN-HL	IN-TL	IN-NC			

by	IN	IN-HL	IN-TL	RB	NIL	IN-NC	
everyone	PN						

There	EX	RB	EX-HL				
were	BED	BED-NC					
people	NNS	NNS-HL	VB	NNS-NC			
everywhere	RB	NN					
covering	VBG	NN					
the	AT	AT-TL	AT-HL	AT-NC	NIL	AT-TL-HL	
roads	NNS	NNS-HL					
along	IN	RB	RB-HL				
the	AT	AT-TL	AT-HL	AT-NC	NIL	AT-TL-HL	
route	NN	FW-NN					
from	IN	IN-HL	IN-TL				
the	AT	AT-TL	AT-HL	AT-NC	NIL	AT-TL-HL	
BJP	NOT_FOUND						
headquarters	NN	NNS					
to	TO	IN	IN-HL	TO-HL	IN-TL	NPS	NIL
	QL	TO-NC	IN-NC				
the	AT	AT-TL	AT-HL	AT-NC	NIL	AT-TL-HL	
Smriti	NOT_FOUN						

	D					
Sthal	NOT_FOUND					
from	IN	IN-HL	IN-TL			
side	NN	NN-HL	JJ			
to	TO	IN	IN-HL	TO-HL	IN-TL	NPS	NIL
	QL	TO-NC	IN-NC			
side	NN	NN-HL	JJ			
with	IN	IN-HL	RB	IN-NC		
security	NN					
personnel	NNS	NNS-HL	FW-JJ			
maintaining	VBG					
strict	JJ					
vigil	NN					
to	TO	IN	IN-HL	TO-HL	IN-TL	NPS	NIL
	QL	TO-NC	IN-NC			
ensure	VB					
that	CS	WPS	DT	QL	WPO	CS-HL	DT-NC
	NIL	WPS-NC	WPO-NC	CS-NC	WPS-HL	
nothing	PN					
goes	VBZ					
wrong	JJ	JJ-HL	RB	NN	JJ-NC	

Rules for Development Data:

The following rules can be elaborated as <TAG1> changes to <TAG2> if <CONDITION> or <IN_CONTEXT_OF TAG3>

1. NOT_FOUND NP CURENT-WORD-IS-CAP YES
2. NN VB PREV-TAG DOZ
3. VBN VBD PREV-TAG NOT HVZ OR HVD
4. WDT WPS NEXT-TAG DO
5. TO IN NEXT-TAG AT OR NN

English Testing Data

During	IN	IN-TL						
a	AT	AT-HL	FW-IN	NIL	AT-NC	NN	FW-IN-TL	AT-TL
visit	NN	VB						
to	TO	IN	IN-HL	TO-HL	IN-TL	NPS	NIL	QL
	TO-NC	IN-NC						
the	AT	AT-TL	AT-HL	AT-NC	NIL	AT-TL-HL		
Cleveland	NP-HL	NP						
Indians	NPS	NPS-HL	NPS-TL					
Beane	NOT_FOUND							
meets	VBZ-HL	VBZ	NNS					
Peter	NP	NP-TL						
Brand	NN	QL						
a	AT	AT-HL	FW-IN	NIL	AT-NC	NN	FW-IN-TL	AT-TL
young	JJ	NN	JJ-NC					

Yale	NP-TL	NP					
economics	NN						
graduate	NN	VB	JJ				
with	IN	IN-HL	RB	IN-NC			
radical	JJ	NN					
ideas	NNS						
about	IN	RB	RP	IN-HL	NIL		
how	QL	WRB	WQL	WRB-HL	WRB-NC		
to	TO	IN	IN-HL	TO-HL	IN-TL	NPS	NIL	QL
	TO-NC	IN-NC					
assess	VB						
player	NN						
value	NN	NN-HL	VB				

Beane	NOT_FOUND					
tests	NNS	NNS-HL	VBZ			
Brand's	NOT_FOUND					
theory	NN	NN-HL				
by	IN	IN-HL	IN-TL	RB	NIL	IN-NC
asking	VBG					
whether	CS					
he	PPS	PPS-N	NIL			

	C					
would	MD	MD-HL				
have	HV	HV-HL	HV-NC			
drafted	VBD	VBN				
Beane	NOT_FOUND					
out	RP	IN	PP$	RP-HL	IN-HL	RB-NC	RP-NC
of	IN	IN-TL	IN-HL	IN-NC	IN-TL-HL	NIL
high	JJ	NN	JJ-HL	RB		
school	NN	NN-HL	VB	NN-TL		

Though	CS						
scouts	NNS						
considered	VBD	VBN					
Beane	NOT_FOUND						
hugely	RB						
promising	JJ	VBG					
his	PP$	PP$$	NIL	PP$-NC	PP$-TL		
career	NN						
in	IN	RP	IN-HL	IN-TL	IN-NC	RP-HL	NN	NIL
	RP-NC	FW-IN					
the	AT	AT-TL	AT-HL	AT-NC	NIL	AT-TL-HL	

Major	NN-TL	NP	JJ	JJ-TL				
Leagues	NNS							
was	BEDZ	BEDZ-HL	BEDZ-NC					
disappointing	JJ	VBG						

Brand	NN	QL						
admits	VBZ							
that	CS	WPS	DT	QL	WPO	CS-HL	DT-NC	NIL
	WPS-NC	WPO-NC	CS-NC	WPS-HL				
based	VBN	VBD						
on	IN	IN-TL	RP	IN-HL	NIL	IN-NC		
his	PP$	PP$$	NIL	PP$-NC	PP$-TL			
method	NN	NN-HL						
of	IN	IN-TL	IN-HL	IN-NC	IN-TL-HL	NIL		
assessing	VBG							
player	NN							
value	NN	NN-HL	VB					
he	PPS	PPS-NC	NIL					
would	MD	MD-HL						
not	-HL	*-NC						
have	HV	HV-HL	HV-NC					

drafted	VBD	VBN				
him	PPO	PPO-NC				
until	CS	IN				
the	AT	AT-TL	AT-HL	AT-NC	NIL	AT-TL-HL
ninth	OD					
round	NN	VB	JJ	RB	IN	

Impressed	VBD	VBN				
Beane	NOT_FOUND					
hires	VBZ					
Brand	NN	QL				
as	CS	QL	CS-HL	IN	CS-TL	RB	NIL
his	PP$	PP$$	NIL	PP$-NC	PP$-TL	
assistant	NN	JJ				
manager	NN	NN-HL				

Addition rules for accommodating Testing Data:

1. JJ VBG PREV-TAG RB OR BEDZ
2. QL WRB NEXT-1-OR-2-TAG VB
3. NNS VBZ NEXT-TAG NP
4. RP IN NEXT-1-OR-2-OR-3-TAG NN
5. CS IN NEXT-1-OR-2-OR-3-TAG NN

We see that after applying the above rules, all the cases of ambiguity in both the development data and testing data were resolved.

Task 2:

Hindi Development Data

जब	RB	CC	PRL	NN
मिठाईवाले	NN			
ने	PSP			
किसान	NN			
को	PSP			
मजे	JJ			
से	PSP	RB	VM	
उसकी	PRP			
दुकान	NN			
की	PSP			
मिठाइयों	NN			
की	PSP			
खूशबू	NN			
का	PSP			
आनंद	NN			
लेते	VM			
देखा	VM			
तब	CC	JJ	RB	PRL
उससे	RB			
किसान	NN			

की	PSP			
खुशी	NN	RB		
देखी	VM			
नहीं	NEG			
गई	JJ	VAUX		

वह	PRP	RB	JJ		
किसान	NN				
के	PSP				
पास	JJ	RB	NN	PSP	
गया	VM	JJ			
और	CC	JJ	PR	INTF	QTF
बोला	VM				
पैसे	NN				
निकालो	VM	NN			

किसान	NN				
हैरान	JJ				
हुआ	VAUX	JJ			
और	CC	JJ	PR	INTF	QTF
बोला	VM				
कि	PSP				
मैंने	PRF				
तो	RPD	CC	RB		

मिठाई	NN				
नहीं	NEG				
खरीदी	VM				
और	CC	JJ	PR	INTF	QTF
न	NEG				
ही	RPD	RB			
चखी	VM				
है	VAUX				
फिर	CC	VM			
पैसे	NN				
किस	VM	PRQ	DMQ		
बात	NN				
के	PSP				

हलवाई	NN		
बोला	VM		
भले	RB	JJ	CC
ही	RPD	RB	
तुमने	PRP		
मिठाई	NN		
नहीं	NEG		
ली	VM		
हो	VAUX		

लेकिन	CC	RB	
मेरी	PRF		
बनाई	VM		
मिठाई	NN		
की	PSP		
खुशबू	NN		
का	PSP		
आनंद	NN		
तो	RPD	CC	RB
लिया	VM		
है	VAUX		

Rules for Development Data:

The following rules can be elaborated as <TAG1> changes to <TAG2> if <CONDITION> or <IN_CONTEXT_OF TAG3>

1. RB PSP NEXT-1-OR-2-TAG NN
2. JJ RB NEXT-TAG VM
3. JJ VAUX PREV-1-OR-2-TAG VM
4. VM PRQ NEXT-TAG NN
5. RB CC NEXT-TAG NOT VM

Hindi Testing Data

एक	QTC	JJ	NN
किसान	NN		
था	VM	VAUX	

_____

वह	PRP	RB	JJ
अपने	PRP	PRF	
खेतों	NN		
में	PSP		
काम	NN		
कर	VM	RB	NN
घर	NN		
लौट	VM	NN	
रहा	VAUX		
था	VAUX		

रास्ते	NN		
में	PSP		
ही	RPD	RB	
एक	QTC	JJ	NN
हलवाई	NN		
की	PSP		
दुकान	NN		
थी	VM	VAUX	

उस	DM	PRL			
दिन	NN				
किसान	NN				

Word	Col1	Col2	Col3	Col4	Col5
ने	PSP				
कुछ	QTF	RB	PRQ	INTF	JJ
ज्यादा	JJ	RB	INTF	QTF	
काम	NN				
कर	VM	RB	NN		
लिया	VAUX				
था	VAUX				
और	CC				
उसे	PRP				
भूख	NN				
भी	RB				
बहुत	JJ	RB	INTF	QTF	
लग	VM	VAUX			
रही	VAUX				
थी	VAUX				

Word	Col1	Col2	Col3	Col4
ऐसे	PRP	DMD		
में	PSP			
जब	RB	CC	PRP	NN
वह	PRP	RB	JJ	
हलवाई	NN			
की	PSP			
दुकान	NN			
के	PSP			

पास	JJ	RB	**NN**	PSP
से	**PSP**	RB	VM	
गुजरा	**VM**			
तो	RPD	**CC**	RB	
उसे	**PRL**			
मिठाइयों	**NN**			
की	**PSP**			
खुशबू	**NN**			
आने	**VM**			
लगी	**VAUX**			

वह	**PRP**	RB	JJ
वहां	**PRP**	RB	JJ
खुद	**PRF**	RB	
को	**PSP**		
रोके	**VM**		
बिना	**NEG**	RB	PSP
नहीं	**NEG**		
रह	**VM**	VAUX	
पाया	**VAUX**	NN	
लेकिन	**CC**	RB	
उस	**DMD**	PRP	
दिन	**NN**		
उसके	**PRP**		

पास	JJ	RB	NN	PSP
ज्यादा	JJ	RB	INTF	QTF
पैसे	NN			
नहीं	NEG			
थे	VAUX			

ऐसे	PRP	DMD	
में	PSP		
वह	PRP	RB	JJ
मिठाई	NN		
खरीद	VM	NN	
नहीं	NEG		
सकता	VAUX		
था	VAUX		

तब	CC	JJ	RB	PRL	
वह	PRP	RB	JJ		
कुछ	QTF	RB	PR	INTF	JJ
देर	NN	JJ	RB		
वहीं	PRP	RB			
खड़े	VM				
होकर	VAUX				
मिठाइयों	NN				
की	PSP				

सुगंध	NN				
का	PSP				
आनंद	NN				
लेने	VM				
लगा	VAUX				

Additional rules for Testing Data:

1. JJ QTF NEXT-1-OR-2-TAG NN
2. JJ NN NEXT-TAG PSP
3. JJ NN PREV-TAG PRP

We see that after applying the above rules, all the cases of ambiguity in both the development data and testing data were resolved.