CS6320, Fall 2019
SwapnaChintapalli – SXC180048

HomeWork 2

1. NFSA to Regular Expression

a. (aa?b)+
b. a(b|ba)(aba?)*

2. Bigram Probabilities:

a. No Smoothing

P(standard|the) = 0.0008161044613710555
P(turbo|standard) = 0.2
P(engine|turbo) = 0
P(is|engine) = 0
P(hard|is) = 0
P(to|hard) = 0.75
P(work|to) = 0.004513217279174726

P(Input_Sentence) = P(standard|the) * P(turbo|standard) * P(engine|turbo) * P(is|engine) * P(hard|is) * P(to|hard) * P(work|to)
= 0

b. Add-one Smoothing

P(standard|The) = 0.0003546728143287817
P(turbo|standard) = 0.00039411455596426696
P(engine|turbo) = 0.0001315097317201473
P(is|engine) = 0.000131250820317627
P(hard|is) = 0.00012423903590508137
P(to|hard) = 0.0005259006047856955
P(work|to) = 0.0008740303725554463

P(Input_Sentence) = P(standard|the) * P(turbo|standard) * P(engine|turbo) * P(is|engine) * P(hard|is) * P(to|hard) * P(work|to)
= 1.377839280853547e-25

c. Good-Turing Discounting based Smoothing

a. Note: Please the set the Good Turing smoothed count C*_(c)=0 if N_(c+1)=0

P(standard|the) = 4.520905085529811e-05
P(turbo|standard) = 1.2824430746649818e-05

P(engine|turbo) = 0.2483575627283251
P(is|engine) = 0.2483575627283251
P(hard|is) = 0.2483575627283251
P(to|hard) = 4.520905085529811e-05
P(work|to) = 0.00010794966406900566
P(Input_Sentence) = P(standard|the) * P(turbo|standard) * P(engine|turbo) * P(is|engine) * P(hard|is) * P(to|hard) * P(work|to)
= 4.334533243595371e-20

3. Transformation Based POS Tagging (40 points)

*The*_DT *standard*_?? *Turbo*_NN *engine*_NN *is*_VBZ *hard*_JJ *to*_TO *work*_??

A: Sol:

Most Probable tags:

Standard: NN
Work: VB

Top rules generated using Brill's transformation – based POS tagging algorithm are:

From	To	Condition	Score
VBP	VB	MD	54
NN	VB	MD	46
VBD	VBN	VBZ	38
VBN	VBD	PRP	35
VB	VBP	PRP	32
POS	VBZ	PRP	31
VB	VBP	NNS	30
VBP	VB	TO	28
NN	VB	TO	27
VBN	VBD	NNP	27
VBD	VBN	VB	22

After applying rules, final solution generated is standard_NN, work_VB

Final Solution:

The_DT standard_NN Turbo_NN engine_NN is_VBZ hard_JJ to_TO work_VB

B: Sol:

*The*_DT *standard*_?? *Turbo*_NN *engine*_NN *is*_VBZ *hard*_JJ *to*_TO *work*_??

Sol:

1. Standard: Word given tags: NN, JJ
   NN = P(standard/NN) * P(NN/DT) * P(NN/NN) = 0.0006305170239596469 * 0.5094639732781931 * 1 = **0.0003212257082460234**

   JJ = P(standard/JJ) * P(JJ/DT) * P(NN/JJ) = 0.000774593338497289 * 0.22808970892317482 * 0.5435063258455978 = 9.60249416421585e-05

2. Work: Word given Tags: VB, VBP, NN
   VB = P(work/VB) * P(VB/TO) = 0.0035035035035035035 *0.63571889103804 = **0.002227243361995135**

   VBP = P(work/VBP) * P(VBP/TO) = 0.011400651465798045 * 0 = 0

   NN = P(work/NN) * P(NN/TO) = 0.00234192037470726 * 0.03288201160541586 = 7.700705294008397e-05

Final Solution:

The_DT standard_NN Turbo_NN engine_NN is_VBZ hard_JJ to_TO work_VB