Name: _____ Reg #: _____ Section: _____

National University of Computer and Emerging Sciences, Lahore Campus

	Course:	**Natural Language Processing**	**Course Code:**	**CS 535**
	Program:	**MS(Computer Science)**	**Semester:**	**Spring 2019**
	Duration:	**30 Minutes**	**Total Marks:**	**10**
	Paper Date:	**15-Feb-19**	**Weight**	**5%**
	Section:	**CS**	**Page(s):**	**2**
	Exam:	**Quiz 1**		

Q1) For the following questions, assume we are using a corpus completely summarized by the unigram counts below (thus V = 11):

Unigram counts:

red 29	retrieval 41	tree 33
apple 34	leaf 1	singing 12
orange 18	skin 4	table 9
animal 1	material 49	

N = Total words = 231

a) What are the following probabilities:

P_{MLE}(orange) = 18/231 = 0.077

P_{MLE}(skin) = 4/231 = 0.017

P_{MLE}(boot) = 0

b) Now assume we are using Laplace smoothing. What are the following probabilities?

$P_{Laplace}$(boot) = 1/242 = 0.00413

$P_{Laplace}$(skin) = 5/242 = 0.0206

Name: _____ Reg #: _____ Section: _____

Q2) You are given the following training corpus: [5 + 5 = 10 Marks]

<s> I like cars </s>
<s> cars like I </s>
<s> We like bikes </s>
<s> I do not like bikes and cars </s>

a) Calculate the probability of following test sentence. Include <s> and </s> in your counts just like any other token. λ_1 = trigram weight, λ_2 = bigram weight, λ_3 = unigram weight, λ_1 = 0.5, λ_2 = 0.3, λ_3 = 0.2

<s> I like bikes </s>

i. Unigram Model

P (<s> I like bikes </s>) = P(I) * P(like) * P(bikes) = 4/24*3/24 *4/24*2/24*4/24
= 0.000048

ii. Bigram Model

P (<s> I like bikes </s>) = P(I | <s>) * P(like | I) * P(bikes | like) * P(</s> | bikes)
= 2/4*1/3 *2/4*1/2 =0.0412

iii. Trigram Model

P (<s> I like bikes </s>) = P(I | <s> <s>)* P(like | <s> I) * P(bikes |I like) * P(</s> | like bikes)
= 2/4*1/2 *0/1*1/2 = 0

iv. **Trigram language model with linear interpolation**.

P (<s> I like bikes </s>) = $P_{interpolated}$(I | <s> <s>)* $P_{interpolated}$(like | <s> I) * $P_{interpolated}$ (bikes |I like) * $P_{interpolated}$ (</s> | like bikes)

$P_{interpolated}$(I | <s> <s>) = 0.5 * P(I | <s> <s>) + 0.3* P(I | <s>) + 0.2* P(I) = 0.5*(2/4) + 0.3*(2/4) + 0.2*(4/24)=0.43
$P_{interpolated}$(like |<s> I)= 0.5 * P(like | <s> I) + 0.3* P(like | I) + 0.2* P(like) = 0.5*(1/2) + 0.3*(1/3) + 0.2*(4/24)=0.38

Name: _____ Reg #: _____ Section: _____

$P_{interpolated}$(bikes |I like)= 0.5 * P(bikes |I like) + 0.3* P(bikes | like) + 0.2* P(bikes) = 0.5*(0/1) + 0.3*(1/4) + 0.2*(2/24)=0.092

$P_{interpolated}$(</s> | like bikes)= 0.5 * P(</s> | like bikes) + 0.3* P(</s> | bikes) + 0.2* P(</s>) = 0.5*(1/2) + 0.3*(1/2) + 0.2*(4/24)=0.43

P (<s> I like bikes </s>) = $P_{interpolated}$(I | <s> <s>)* $P_{interpolated}$(like | <s> I) * $P_{interpolated}$ (bikes |I like) * $P_{interpolated}$ (</s> | like bikes) = 0.43*0.38*0.092*0.43 = 0.0064

b) Calculate the probability of P(cars | like) using Kneser Ney smoothing from the corpus given above. d = discounting factor = 0.5

P(cars | like) = (1 – 0.5) / 4 + 3*(0.5/4)* (3/18) = 0.187

Continuation count:
<s> 0
I 2
like 4
cars 3
we 1
</s> 4
Bikes 1
Do 1
Not 1
and 1

Continuation count of all words = 18