Homework 8

N-gram exercise

Exercise

Given the Corpus:

- He is a nice guy
- It's nice to get your letter
- I am going to meet you
- I would like a meat sandwich

Estimate the probability of the following two sentences (with smoothing):

- Nice to meet you
- Nice to meat you

Hint

Add Start & End Tag:

- <s> He is a nice guy <e>
- <s> It's nice to get your letter <e>
- ...

The bigrams:

- <s> He, He is, is a, a nice, nice guy, guy <e>
- <s> It's, It's nice, nice to, to get, get your, your letter, letter <e>
- ...

Then the probabilities (the same as Exercise 1 in tutorial hour) of the 2 sentences both become $0 \rightarrow \text{Apply smoothing}$

The simplest way to do smoothing is to add 1 to all the bigram counts C before normalization.

e.g.

	nice	sandwich	meat	going	letter	guy	1	meet	like	your	he	to	am	а	get	would	it's	is	you	<e></e>
i-1 <s></s>							2				1						1			
nice						1	_				_	1					_			
sandwich																				1
meat		1																		
going												1								
letter																				1
guy																				1
1													1			1				
meet																			1	
like														1						
your					1															
he																		1		
to								1							1					
am				1																
a	1		1																	
get										1										
would									1											
it's	1																			
is														1						
you																				1



j-1	nice	sandwich	meat	going	letter	guy	T	meet	like	your	he	to	am	а	get	would	it's	is	you	<e></e>
<s></s>	1	1	1	1	1	1	3	1	1	1	2	1	1	1	1	1	2	1	1	1
nice	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1
sandwich	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
meat	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
going	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1
letter	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
guy	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	1	1
meet	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
like	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1
your	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
he	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
to	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1
am	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
а	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
get	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
would	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
it's	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
is	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1
you	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2