IHLT Exam – 2020 (Start time 11:30; End time 14:30)

1. (4 point) Given the following proposals (G1, G2 and G3) for probabilistic context free grammars (PCFG) with start symbol NP,

G1:
[r1] NP → NNS PP (0.3)
[r2] NP → DT NN (0.6)
[r3] NP → NN (0.1)
[r4] PP → IN NP (1.0)
[r5] DT → the (1.0)
[r6] IN → in (0.3)
[r7] IN → with (0.7)
[r8] NN → loft (0.6)
[r9] NN → money (0.4)
[r10] NNS → boxes (1.0)

G2:
[r1] NP → NP PP (0.7)
[r2] NP → the NN (0.2)
[r3] NP → NN (0.1)
[r4] NP → NNS (0.2)
[r5] PP → IN NP (1.0)
[r6] IN → in (0.4)
[r7] IN → with (0.6)
[r8] NN → loft (0.3)
[r9] NN → money (0.7)
[r10] NNS → boxes (1.0)

G3:
[r1] NP → NP PP (0.4)
[r2] NP → DT NN (0.3)
[r3] NP → NNS (0.15)
[r4] NP → NN (0.15)
[r5] PP → IN NP (1.0)
[r6] DT → the (1.0)
[r7] IN → in (0.5)
[r8] IN → with (0.5)
[r9] NN → loft (0.4)
[r10] NN → money (0.6)
[r11] NNS → boxes (1.0)

a) Answer the following questions. Justify your answers (answers without justification or with wrong ones will be considered wrong answers).

   1. Which of the 3 proposals are correct PCFGs?
   2. Which of the 3 proposals are in Chomsky Normal Form (CNF)?
   3. Which of the 3 proposals can be transformed into CNF?
   4. Which of the 3 proposals subsume sentence *"boxes with money in the loft"?*

b) Transform into CNF the correct, transformable proposals that subsume sentence *"boxes with money in the loft"*

c) Draw the best parse trees when apply the PCFGs in CNF resulting from (b) to sentence *"boxes with money in the loft"*. **No algorithm is required**. Provide each node with its respective likelihood.

2. (2 points) Suppose that the following paragraph is part of a training corpus useful for learning mention-pair coreference resolution models. Mentions (identifiers are preceding numbers) and coreference chains (identifiers are subindexes) have been manually annotated. Answer the following points (answers without justifications or with wrong ones will be considered wrong answers).

> 1:[Mark Johns]$_1$ was used to visiting 2:[3:[his]$_1$ grandma]$_2$. However, 4:[that day]$_3$, 5:[Mrs Johns]$_2$ fell ill with 6:[COVID]$_4$. At 7:[9:00 a.m.]$_5$, 8:[the disease]$_4$ forced 9:[her]$_2$ to go to hospital and 10:[he]$_1$ was not able to find 11:[the name of 12:[the hospital]$_7$]$_6$ in 13:[which]$_7$ 14:[15:[his]$_1$ grandma]$_2$ was admitted.

a) From the lists below, which options consist only of suitable positive examples for learning a model based on closest-first strategy? For the rest of options, justify the reason by which you think they consist of at least one unsuitable positive example.

1.- (1,3) (5,9) 2.- (2,5) (9,14) 3.- (2,14) (3,10) 4.- (10,15) (12, 13)

b) From the list below, which options are suitable negative examples for learning a model based on closest-first strategy? For the rest of options, justify the reason by which you think they are unsuitable negative examples.

1.- (5,8) 2.- (3,4) 3.- (10,15) 4.- (8,14)

c) Are Hidden Markov Models (HMMs) a good approach for learning a mention-pair model? Justify the answer

d) Given the following non-zero probabilities provided by a mention-pair model, apply best-first strategy to the paragraph and answer the coreference chain achieved that includes mention 1:[Mark Johns]. Consider a probability threshold of 0.7. Justify your answer.

P(1,3)=0.71
P(3,4)=0.45
P(1,5)=0.88; P(2,5)=0.90; P(3,5)=0.66;
P(1,8)=0.51; P(4,8)=0.73; P5,8)=0.44; P(6,8)=0.79
P(1,9)=0.72; P(2,9)=0.87; P(5,9)=0.88; P(6,9)=0.86
P(1,10)=0.78; P(2,10)=0.65; P(3,10)=0.92; P(5,10)=0.55; P(6,10)=0.73; P(9,10)=0.34
P(1,11)=0.69; P(2,11)=0.36; P(3,11)=0.40; P(5,11)=0.67; P(6,11)=0.77; P(8,11)=0.54
P(1,12)=0.77; P(2,12)=0.27; P(3,12)=0.45; P(5,12)=0.59; P(6,12)=0.80; P(8,12)=0.63
P(11,13)=0.91; P(12,13)=0.97
P(1,14)=0.69; P(2,14)=1.0; P(5,14)=0.82; P(6,14)=0.81; P(9,14)=0.87; P(10,14)=0.36
P(1,15)=0.71; P(2,15)=0,60; P(3,15)=1.0; P(9,15)=0.66; P(10,15)=0.73

3. (4 points) Given a HMM model represented by the following tables:

A	DT	JJ	NN	VBP	RB	IN
DT		0.2	0.5	0.2	0.1	
JJ		0.4	0.6			
NN			0.2	0.5	0.1	0.2
VBP	0.2		0.3		0.1	0.4
RB	0.3	0.1	0.5			0.1
IN	0.2	0.2	0.6			

Π	
DT	0.5
JJ	
NN	0.3
VBP	0.2
RB	
IN	

B	stay	back	car	the	in	of
DT				1.0		
JJ		1.0				
NN	0.4	0.1	0.5			
VBP	0.6	0.4				
RB		1.0				
IN					0.5	0.5

DT: determiner; JJ: adjective; NN: singular common noun; VBP: verb; RB: adverb; IN: preposition

a) Apply the model to disambiguate the following morphological analysis. Provide clearly the resulting dynamic table, the resulting POS-tags sequence and its likelihood.

stay	*in*	*the*	*back*	*of*	*the*	*back*	*car*
NN	IN	DT	NN	IN	DT	NN	NN
VBP			VBP			VBP	
			JJ			JJ	
			RB			RB	

b) Is the resulting POS-tags sequence really the correct one? Justify the answer in up to two lines (Answer without justification or with a wrong one will be considered wrong answer). How could the HMM model be improved?