Section 1

a)

i
Taking a rolling average of the gradients has a smoothing effect as it consists of weighted average of the past gradients, with weights reducing for old time steps. Thus, the gradient calculated is a long term one, and is immune to fluctuations in current gradients.

ii
It helps to normalise the parameter updates, as it reduces the effective learning rate of parameters with high gradients and increases the effective learning rate of parameters with low gradients.

b)

i
$\gamma = 1/p_drop$
During training p_drop ratio of the neurons are dropped, but during testing all the neurons are present.
So the output is scaled by γ so that the future layers receive values in the same scale as it saw during training.

ii
Dropout during training helps in regularisation, as it helps the future layers to use as many neurons in making decisions, since either of them could be dropped randomly.
During test time we want our results to be deterministic across multiple forward passes and hence we cannot induce randomness into the model.

Section 2

a)

Stack	Buffer	New dependency	Transition	step
[ROOT]	[I, parsed, this, sentence, correctly]		Initial Configuration	0
[ROOT, I]	[parsed, this, sentence, correctly]		Shift	1
[ROOT, I, parsed]	[this, sentence, correctly]		Shift	2
[ROOT, parsed]	[this, sentence, correctly]	I <- parsed	Left-Arc	3
[ROOT, parsed, this]	[sentence, correctly]		Shift	4

Stack	Buffer	New dependency	Transition	step
[ROOT, parsed, this, sentence]	[correctly]		Shift	5
[ROOT, parsed, sentence]	[correctly]	this <- sentence	Left-Arc	6
[ROOT, parsed]	[correctly]	parsed -> sentence	Right-Arc	7
[ROOT, parsed, correctly]	[]		Shift	8
[ROOT, parsed]	[]	parsed -> correctly	Right-Arc	9
[ROOT]	[]	Root -> parsed	Right-Arc	10

b)

At every step, a word is either pushed into the stack or popped off it.
Since there are n words and every word goes on and off the stack once, there are a total number of $2n$ steps

f)

i
Error: Verb phrase attachment error
Incorrect dependency: wedding -> fearing
Correct dependency: heading -> fearing

ii
Error: Coordination attachment error
Incorrect dependency: makes -> rescue
Correct dependency: rush -> rescue

ii
Error: Prepositional phrase attachment error
Incorrect dependency: named -> midland
Correct dependency: guy -> midland

ii
Error: Modifier attachment error
Incorrect dependency: elements -> most
Correct dependency: crucial -> most