Assignment 3: Dependency Parsing

1. (a) (i)
m is consisted of a proportion of the current gradient (usually a1) and a proportion of the previous m (usually 0.9). Thus, m would not vary a lot after each update, so as the parameters. This low variance may be helpful to learning since it can help us converge easier to the optimum. It also has the potential to prevent us from overshooting

(ii)
Parameters for the initial steps get larger updates since $\sqrt{v}$ is fairly small at the beginning. This trick mimics the adaptive learning rate, which starts fast and decreases across timesteps.

(b) (i)

$$E_{p_{drop}}[h_{drop}]_i$$

$$= \gamma \cdot p_{drop} \cdot 0 + \gamma \cdot (1 - p_{drop}) \cdot h_i$$

$$= h_i$$

$$\therefore \gamma = \frac{1}{1 - p_{drop}}$$

(ii) We don't want randomness in our prediction results.

2. (a)

Stack	Buffer	New dependency	Transition
[ROOT, parsed, this]	[sentence, correctly]		SHIFT
[ROOT, parsed, this, sentence]	[correctly]		SHIFT
[ROOT, parsed, sentence]	[correctly]	sentence → this	LEFT-ARC
[ROOT, parsed]	[correctly]	parsed → sentence	RIGHT-ARC
[ROOT, parsed, correctly]			SHIFT
[ROOT, parsed]		parsed → correctly	RIGHT-ARC
[ROOT]		ROOT → parsed	RIGHT-ARC

(b) 2n+1 steps. n steps for moving words from buffer to stack and the other n steps for finding out dependencies between words. One extra step for initialization.

(f) (i) Error type: Verb Phrase Attachment Error
Incorrect dependency: wedding → fearing
Correct dependency: I → fearing

(ii) Error type: Coordination Attachment Error
Incorrect dependency: and → rescue
Correct dependency: out → rescue

(iii) Error type: Prepositional Phrase Attachment Error
Incorrect dependency: named → Midland
correct dependency: guy → Midland

(iv) Error type: Modifier Attachment Error
Incorrect dependency: most ⟶ elements
Correct dependency: most ⟶ crucial